Theory, Simulation, and Computation Directorate

Science Highlights

2012

**Los Alamos**
NATIONAL LABORATORY
EST. 1943

Los Alamos National Laboratory

About the cover:
The background image used for the cover has been artistically altered; it is a graphic selected from an article within this document. For further information please refer to page 43, "Mixing in Converging Flows."

Back Cover:
The five images on the back have been selected from various articles throughout this publication. They are a representation of the breadth of science and talent that the Theory, Simulation, and Computational Directorate possesses. Please see articles on pages 7, 32, 43, 47, and 117 to find out more about these images.
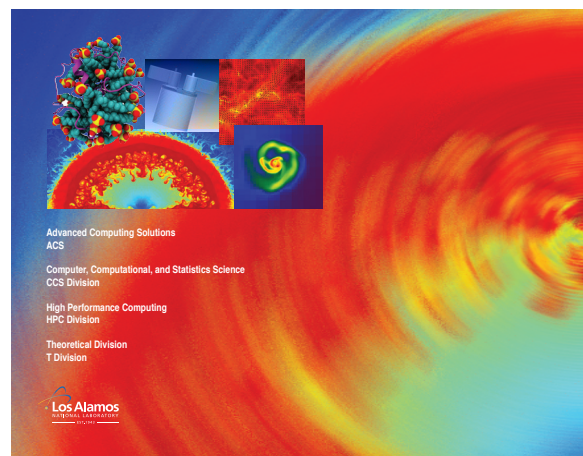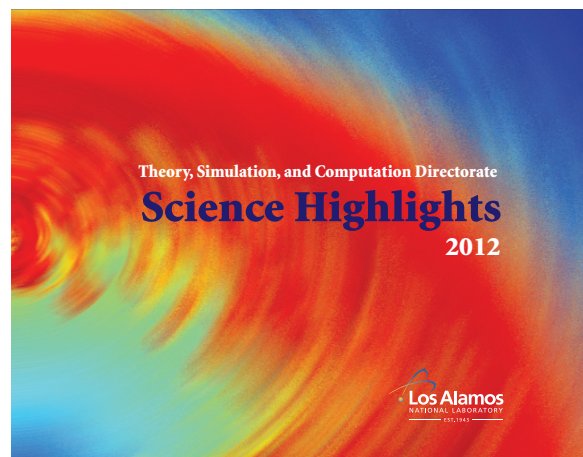
ADTSC Science Highlights 2012

Associate Directorate for Theory, Simulation, and Computation

Preface

The Theory, Simulation, and Computation Directorate encompasses a very broad array of technical disciplines. However, our goal is to integrate across these disciplines and work with our experimental colleagues to develop a truly predictive understanding of the complex problems the country faces in national security. Combining validated theory, algorithmic developments, and high-performance computing, we can raise the notion of the scientific method to Information Science and Technology for Prediction.

This annual publication presents representative examples of the interdisciplinary research that we undertake in ADTSC, partnered with our colleagues from across the Laboratory, and indeed from across the globe, toward "Integrating Information, Science, and Technology for Prediction."

Paul J. Dotson
Associate Director (acting), Theory, Simulation, and Computation Directorate

Paul J. Dotson
Associate Director (acting)

Audrey L. Archuleta
ADTSC Chief of Staff

Stephen R. Lee
CCS Division Leader

John F. Morrison
HPC Division Leader

Antonio Redondo
T Division Leader

Publications Team

Sharon Mikkelson
Design and Coordination

Kathy Pallis
Writing/Editing

Contents

Applied Mathematics and Fluid Dynamics

Astrophysics and Cosmology

Climate, Atmospheric, and Earth Systems Modeling

Information Science and Technology

Atomic, Nuclear, and High-Energy Physics

Chemistry and Biology

Materials Science

Funding Acknowledgment Cross Reference

Applied Mathematics and Fluid Dynamics

LANL continues its rich scientific tradition of applied and scholarly research contribution to the fields of mathematics, fluid dynamics, and particle transport. LANL's impressive breadth of research speaks to this history, but also ties LANL scientists to the present and future with the significant integration of world-class theory, modeling, computational science, and experimentation through advanced data assimilation and uncertainty quantification. The articles in this section provide a vignette of this work in a variety of energy sectors (nuclear, carbon capture, combustion) with examples of applications to fluid turbulence, fundamental new methods and numerical algorithms in particle transport, electromagnetics, acoustics, flow dynamics applicable to the earth's oceans, and the behavior of shocked fluids. Most of these advances are made within the fabric of complex multi-physics simulations on the largest, most complex, advanced supercomputers on the planet.

A Nonlinear Krylov Accelerator for the Boltzmann k-Eigenvalue Problem

Matthew T. Calef, Erin D. Fichtl, James S. Warsa,
Markus Berndt, Neil N. Carlson, CCS-2

We compare variants of Anderson Mixing with the Jacobian-Free Newton-Krylov and Broyden methods applied to the k-eigenvalue formulation of the linear Boltzmann transport equation. We present evidence that one variant of Anderson Mixing finds solutions in the fewest number of iterations. We examine and strengthen theoretical results of Anderson Mixing applied to linear problems.

The criticality of fissioning systems is typically described by the k-eigenvalue formulation of the linear Boltzmann transport equation. Physically, the largest eigenvalue, generally denoted by k, is the effective neutron multiplication factor that, in the equation, scales the fission production term to achieve a steady-state solution. The corresponding eigenmode describes the neutron flux profile for that steady-state (i.e., critical) system and, when the system is close to criticality, provides useful information about the distribution of the neutron population in space and velocity.

Mathematically, the equation is a standard eigenproblem for which power iteration is well-suited because the eigenmode of interest is most commonly that with the largest magnitude. For the k-eigenvalue problem, each step of a true power iteration incurs a heavy computational cost due to the expense of fully inverting the transport operator, therefore a nonlinear fixed point iteration (FPI) is generally employed in which an approximate inversion of this operator is performed at each iteration. Recently, more sophisticated nonlinear iteration methods than FPI, predominately Jacobian-Free Newton-Krylov (JFNK) , have been applied with great success [1,2].

No broad comparison of the performance of nonlinear solvers applied to the k-eigenvalue problem has been performed. In [1] we present such a comparison, examining the performance of three nonlinear solvers, JFNK, Broyden's Method, and Anderson Mixing, applied to the k-eigenvalue problem. A variant of Anderson Mixing [2], first described in [3], is of particular interest because, in the experience of the authors, it is frequently computationally more efficient than JFNK and Broyden's Method.

JFNK is an inexact Newton's method in which the inversion of the Jacobian is performed to arbitrary precision using a Krylov method (most commonly the generalized minimum residual method, GMRES) and the Jacobian itself is never formed, but rather its action approximated using finite differences of arbitrarily close state data. JFNK can be expected to converge quadratically in a neighborhood containing the solution. Each iteration of JFNK requires a nested inner iteration and the bulk of the computational effort is expended in this inner Krylov inversion of the Jacobian at each outer Newton step. At the end of each inversion, the accumulated Krylov space is discarded even though the Jacobian is expected to change minimally during the final Newton steps when a similar space will be rebuilt in the next Newton iteration. In effect, at the end of each iteration, JFNK discards information that may be of use in successive iterations.

In its standard formulation, Broyden's method uses differences in state from successive iterations to make low-rank updates to the Jacobian. The Sherman-Morrison-Woodbury update rule is then used to compute the action of the inverse of the Jacobian after such an update. While Broyden's method is restricted to low-rank updates, it provides an explicit representation of the Jacobian, allowing one to employ the Dennis-More condition to show that it converges super-linearly in a neighborhood containing the solution.

Anderson Mixing [2] uses differences in state from successive iterations to infer information about the inverse of the Jacobian, which is assumed to be roughly constant in a neighborhood containing all the iterates. Unlike Broyden's method, the updates can be of arbitrary rank. Recent results by Walker and Ni [6] show that, with mild assumptions, Anderson Mixing applied to a linear problem performs as well as GMRES. In this regard, Anderson Mixing may be thought of as a nonlinear version of GMRES. In independent work, Carlson and Miller formulated a so-called nonlinear Krylov acceleration [3] method, which we show to be a variant of Anderson Mixing. We refer to this method as NKA.

Our test problem is a cylinder with a 3.5-cm radius and a height of 9 cm modeled in 2D cylindrical coordinates. The problem consists of a central 5-cm layer of Boron-10 with 1-cm-thick water layers on either side and 1-cm layers of highly enriched uranium on the ends. The top, bottom,
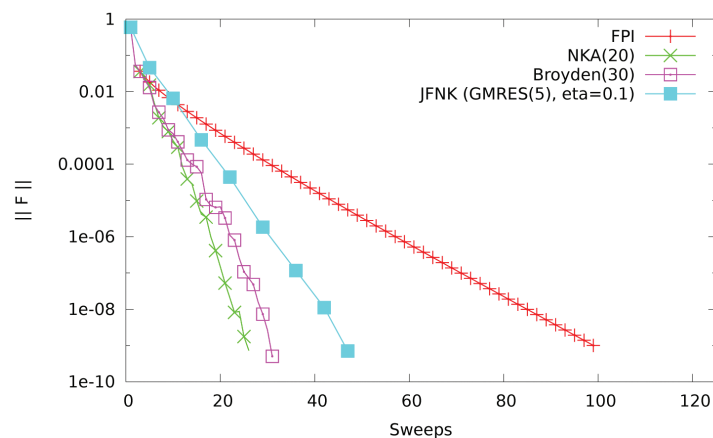
Fig. 1. The PARTISN code applied to the unreflected cylinder problem. Comparison of the performance of four nonlinear solvers. Scaled L2-norm of the residual as a function of number of sweeps for the various methods and subspace sizes. Note that, for JFNK, points plotted on the lines indicate when a JFNK iteration starts (JFNK requires multiple sweeps per iteration). For the other methods, we plot one point per two iterations of the method.

and radial boundaries are either all vacuum or all reflective (the inner radial boundary is a symmetry condition in cylindrical coordinates). This a difficult problem (i.e., one with a high dominance ratio) because the problem is symmetric, having two fissile regions that are effectively decoupled because of the reflector (water) and absorber (boron) between them, which means the second eigenmode is close to the fundamental mode–reflection further increases the difficulty of the problem. A 16-group Hansen-Roach cross-section data set is used to generate the results presented here.

A 175 (r-axis) by 450 (z-axis) mesh of equally sized squares is used in PARTISN [7] for both the reflected and unreflected problems. The Capsaicin results are computed on an unstructured mesh comprising roughly the same number of cells in the r and z axes as the PARTISN computation, for a total of 79,855 (possibly non-convex) polygons, each of which has 3 to 6 vertexes on a cell. While these two codes are both deterministic, discrete ordinates transport codes, they are quite different in the spatial discretizations they employ and in other aspects of implementation, which can cause differences in iterative behavior and result in slightly different solutions.

In Fig. 1, we present a comparison of methods described above for the unreflected cylinder problem run with PARTISN. The particular runs that are compared here represent the best set of parameters that we found for each of the methods. Points plotted on the lines represent two nonlinear iterations of the method, except for JFNK where we plot a point at the beginning of each JFNK inner iteration. We plot the error on the vertical axis, while on the horizontal axis we plot the number of sweeps that, for this class of problems, is an excellent measure of computational cost. The results shown in Fig. 1 are representative of our more detailed study [1]. NKA performs best, with Broyden second, JFNK third; FPI is not competitive. Moreover, NKA exhibited the least sensitivity with respect to its parameters.

The benefit of JFNK is its Newton-like convergence, achieved by developing an arbitrarily accurate approximation of the inverse of the Jacobian at each of these outer iteration, but the cost is repeated function evaluations–or sweeps in our case–in each of these outer iterations and the wasteful discarding of potentially useful subspace information. In contrast, Broyden and Anderson Mixing only perform a single function evaluation at each iteration, but continue to use old information from previous iterations to improve their estimate of the Jacobian. The drawback for these methods is that the approximate Jacobian or its inverse is based on an amalgam of new and old information, so they are unlikely to converge in fewer iterations than Newton's method. Performance of all these methods will clearly depend on how the Jacobian is changing from iteration to iteration, and how information collected at each function evaluation is used. Memory requirements for Broyden are half that of NKA, making it an attractive alternative.

Our numerical results indicate that Anderson Mixing in the form of NKA found solutions in the PARTISN and Capsaicin codes for both the reflected and unreflected problems in the fewest number of function evaluations and the shortest runtimes. These physics problems represent actual reactor geometries and were run at a large scale. Our results highlight the strength of this method: regularity, consistency, and efficiency. In our results, NKA was shown to bring the norm down to zero smoothly, much as FPI and JFNK do, but with greater efficiency than those methods. Broyden at times achieved excellent performance, did not always demonstrate this same smooth convergence behavior, and often diverged. Based on these results we feel that NKA may be well-suited to other computational physics problems.

[1] Calef, M.T. et al., *J Comput Phys,* submitted (2012).
[2] Anderson, D.G., *J ACM* **12**, 547 (1965).
[3] Carlson, N.N. and K. Miller, *SIAM J Sci Comput* **19**, 728 (1998).
[4] Gill, D.F. et al., *Nucl Sci Eng* **168**, 37 (2011).
[5] Knoll, D.A. et al., *Nucl Sci Eng* **167**, 133 (2011).
[6] Walker, H.F. and P. Ni, *SIAM J Numer Anal* **49**, 1715 (2011).
[7] Alcouffe, R.E. et al., LANL Technical Report LA-UR-05-3925 (2005).

Adaptive Finite Element Methods for Combustion Modeling

David B. Carrington, T-3; Juan Heinrich, University of
New Mexico; Xiuling Wang, Purdue University;
Darrell W. Pepper, University of Nevada

To facilitate predictive combustion modeling, in particular simulation of internal combustion engines, we are developing *hp*-adaptive finite element methods (FEM) and software that provide a high degree of accuracy and solution robustness. We have developed an FEM projection method, a predictor-corrector scheme (PCS), that has excellent capability over all flow regimes, from incompressible to supersonic flow. The system is applicable to Newtonian and non-Newtonian fluids, multi-species flow, and multi-phase (with spray/injection modeling). These algorithms are capable of representing the physics within an engine. We continue to provide software that many use directly for engine simulation or that they may alter with various models such as sophisticated chemical kinetics, different turbulent closure methods, or other fuel injection methods. Currently the software (KIVA) is in worldwide use for research. We are working to deliver a more predictive capability that when put to use by researchers and corporations should help in our understanding of internal combustion engines, providing for greater efficiency.

To facilitate predictive combustion modeling, in particular simulation of internal combustion engines, we are developing methods that provide a high degree of accuracy and robustness of solution. We also continue to improve the physics modeling methods, with addition of appropriate turbulence closure schemes. We have developed a projection method, a predictor-corrector scheme (PCS) that has excellent capability over all flow regimes, from incompressible to supersonic flow. The system is applicable to Newtonian and non-Newtonian fluids, multi-species flow and multi-phase (with spray/injection modeling). This PCS algorithm can also be used for porous media, solidification, or incompressible solid mechanics modeling with use of appropriate constitutive models.

This PCS method, combined with higher-order polynomial approximation for model-dependent physical variables (*p*-adaptive) along with grid enrichment (locally higher grid resolution–*h*-adaptive) and overset grids for actuated and immersed moving parts, will provide for a high order of accuracy and robust solutions in the next generation of KIVA, particularly on complex domains.

We continue to provide software that others may use directly or that they may alter with various component models such as sophisticated chemical kinetics, different turbulent closure methods, or other fuel injection methods. Our current users are worldwide, from individuals and small research institutions to large corporations. We expect to be able to deliver a more predictive and robust modeling capability that, when put to use by these users, should help in our understanding of internal combustion engines. With that understanding, even greater efficiency should result. We expect sufficient changes in efficiency that we anticipate saving 4 million barrels of fuel per day in the US transportation fleet when this work is combined with other research and innovations in the field. Along with reduced fuel use, greater model predictability will help to reduce dangerous combustion products.

Proceeding with the idea in mind that it is better to have algorithms that are more accurate at a given resolution and provide for higher resolution and accuracy only where and when required, we recently began researching the use of adaptive finite element methods (FEM) [1]. Our construction is a Galerkin-type FEM that utilizes conservative momentum and energy transport. The projection scheme uses precise Petrov-Galerkin stabilization and pressure stabilization.

Projection methods are enjoying great popularity in the finite element community for good reason. These methods, combined with higher-order polynomial approximation for model-dependent physical variables (*p*-adaptive) along with grid enrichment (locally higher grid resolution–*h*-adaptive) and overset grids for actuated and immersed moving parts, are providing for highly accurate and robust solutions in the next generation of KIVA, particularly on complex domains. We have been developing this PCS algorithm for KIVA from our *h*-adaptive and *hp*-adaptive FEM research to create a new KIVA combustion code [2-5].

Fig. 1. Density contours and adapted grid for 2D Mach 2.22 through a 15° compression ramp: a) Inviscid flow, b) Viscous flow

a) Inviscid Flow: Density and adapted grid, showing shock

b) Viscous Flow: Density and adapted grid, showing shock, flow recirculation and detached shock

a)

b)
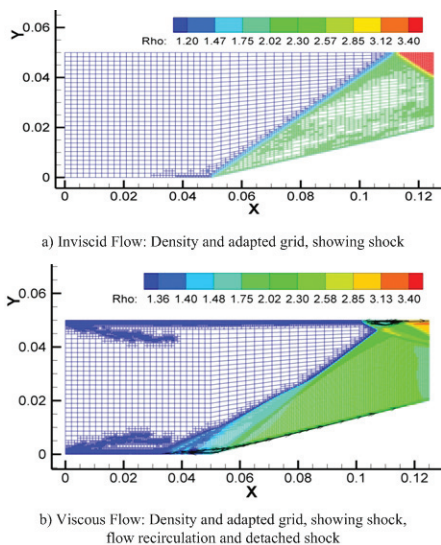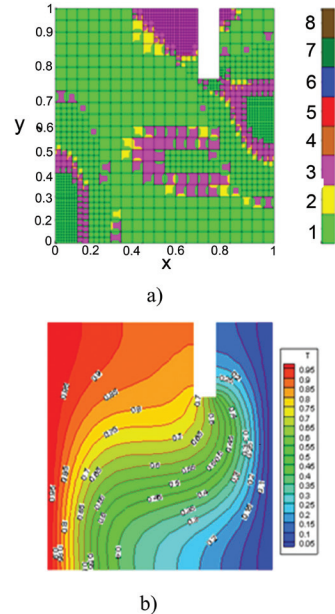
Fig. 2. Partially divided differentially heated enclosure; a) final hp-adaptive mesh, b) isotherms.

a)

b)
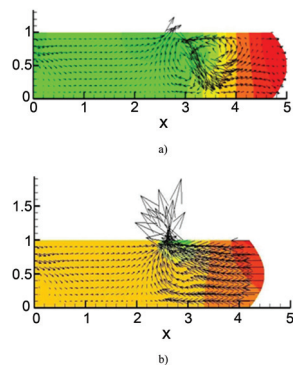
Fig. 3. Immersed moving piston for PCS FEM system; a) and b) showing displacement, velocity vectors (color represents isobars variation).

The immersed actuated parts in the engine are represented by an unstructured overset grid system where local adjustments to the fluid grid occur only when the actuated parts cross the boundary of a fluid element, a robust scheme that works for complex domains. The method maintains second-order spatial accuracy, and is a form of Arbitrary Lagrangian Eulerian (ALE) method, where the fluid grid moves only locally to accommodate a moving overset grid representing the part, such as a valve or piston. Once the part leaves an element, or when the part is outside an element, the element returns to its original shape.

Figure 1 shows the *h*-adaptive solution for both inviscid and viscous Mach 2.22 supersonic flow through a 2D compressor. The inviscid solutions are in exact agreement with the analytical solution, with the viscous solution agreeing with similar solutions [6]. The flow in Fig. 1b includes the viscous terms, and the algorithm is capable of capturing the small zones of recirculation in front to the main shock.

Verification results shown in Fig. 2 are using a PCS fractional-step algorithm [1,4], natural convection in a partially divided enclosure. The left and right walls are maintained at hot and cold temperatures respectively. Figure 2a shows the final *hp*-adaptive mesh where the initial coarse mesh consisted of 388 quadrilateral elements with 435 nodes; the final mesh consists of 1,261 elements and 4,714 degrees of freedom.

In Fig. 3, a 2D piston is shown at two different times during the cycle. The solution is performed for incompressible flow to facilitate development and validation. Error for this formulation against a known solution (a benchmark) is one percent or less depending on resolution; the method shows second-order spatial convergence.

The use of Cubit [http://cubit.sandia.gov/] for engine grid generation is being investigated. Once the grid is implemented in a script language, the grid generation process is automatic. The scripts are often easily altered for various engine domains and grid refinements. This work supports our effort to produce unstructured hexahedral engine grids that are essentially automatic. The file converter from Cubit output has been developed for both KIVA-4mpi and the new FEM formulation. Because the FEM grid is designed for use with the overset grid scheme

discussed previously, it is much easier to produce a quality grid (almost automatically) than it is for the current KIVA-4 using an unstructured grid and the old acutated parts system. The Cubit grids for an airfoil validation problem and for an internal combustion engine are shown in Fig. 4.

Our work continues by: 1) extending the moving parts algorithm to 3D, 2) adding KIVA's injection/spray and chemistry packages to the FEM version, and 3) implementing the method for solving engine simulations using multi-processors (that is, in solutions run in parallel manner).

Fig. 4. Cubit grid generation for FEM KIVA and for KIVA-4 using all hexahedral elements; a) NACA-0012 local mach values from PCS algorithm, b) grid for PCS algorithm, c) grid for KIVA-4.

FEM PCS Grids

a) 3-D NACA Airfoil   b) Engine grid

KIVA-4 Grid

c) Engine grid
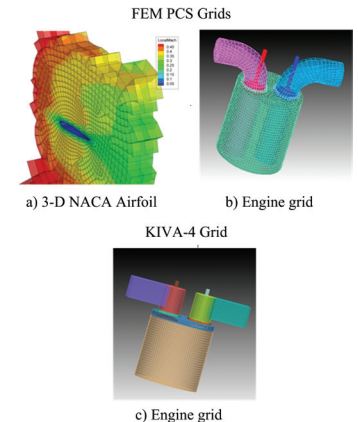
[1] Carrington, D.B., "A Fractional Step hp-adaptive Finite Element Method for Turbulent Reactive Flow," LANL Technical Report LA-UR-11-00466 (2011).
[2] Carrington, D. B. et al., *Comput Model Eng Sci* **61**, 23 (2010).
[3] Wang, X. et al., *Comput Therm Sci* **1**, 121 (2009).
[4] Carrington, D.B., "A Characteristic-based Split hp-adaptive Finite Element Method for Combustion Modeling in KIVA-hpFE," LANL Technical Report LA-UR-09-06527 (2009).
[5] Wang, X. and D. W. Pepper, *J Thermophys Heat Tran* **21**, 190 (2007).
[6] Breuckner, F.P. and Heinrich, J.C., *Int. J. Numer. Meth. Fluids* **32,** 255 (1991).

M-adaptation for the Acoustic Wave Equation

Vitaliy Gyrya, Konstantin Lipnikov, T-5

We have developed a novel adaptive strategy, dubbed m-adaptation, for solving the acoustic wave equation (in the time domain) on square meshes. The m-adaptation is based on selecting the optimal member of a three-parameter Mimetic Finite Difference (MFD) family of second-order schemes. This family contains as particular members the classical methods such as the finite element, finite difference, and a few other more recent methods. The optimal member of the MFD family eliminates the numerical dispersion at the fourth order and the numerical anisotropy at the sixth order. The numerical experiments show that the new approach is consistently better than the classical methods for reducing a long-time integration error.

Fig. 1. A more narrow band of values in the dispersion curves for the m-adaptation method (bottom) compared to the modified quadrature method [2] (top) for various angles θ between the planar wave and the mesh axis for the Courant number $\frac{c\Delta t}{h}=0.75$ indicates smaller anisotropy.
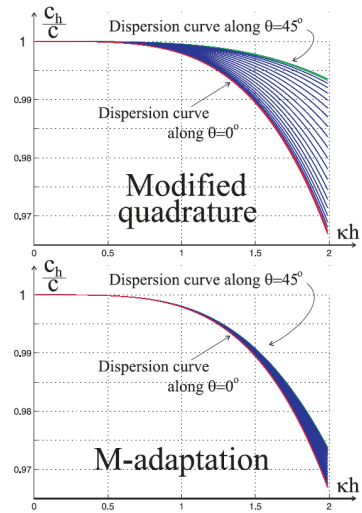
Numerical modeling of wave propagation is essential for a large number of applied problems in acoustics, elasticity, and electromagnetics. The acoustic equation is one of the simplest examples of equation modeling wave propagation. For long integration times, the dominant contributions to an error in the solution come from such numerical artifacts as numerical dispersion and numerical anisotropy. The numerical dispersion is the phenomenon in which the propagation velocity of the wave in the numerical scheme depends on its wavelength, while in the continuum problem there is no such dependence. Typically, the effect of the numerical dispersion is greater on under-resolved waves with ten or fewer points per wavelength, making them travel slower than in the physical problem. As a consequence, the wave does not just arrive at a wrong time (which could be compensated by time rescaling) but it also has a highly distorted profile. The numerical anisotropy is the dependence of the numerical velocity of the wave on its orientation with respect to the mesh. For a 2D acoustic wave equation we developed an adaptation technique, dubbed m-adaptation, that selects an optimal member of a rich parameterized family of second-order methods with smallest (fourth-order) dispersion and (sixth-order) anisotropy.

The semi-discrete form of the acoustic wave equation in the time domain formulation is:

$$Mu_{tt} = Au \qquad (1)$$

where the mass and stiffness matrices M and A are assembled from elemental matrices M_E and A_E.

Since the mass matrix M has to be inverted on every time step, the explicit time discretization of equation (1) is computationally efficient only when the inverse M^{-1} is easy to compute. One of the common approaches is to replace the mass matrix M with a diagonal matrix D by lumping non-diagonal entries to the diagonal. This does not change the order of the numerical scheme but may lead to an undesirable increase of numerical dispersion. Another approach [1] is to replace the inverse M^{-1} with the product $D^{-1}MD^{-1}$, where the inverse is taken only for the diagonal matrix D. Similar to lumping, this approach does not change the order of the numerical scheme but may also result in the increase of the numerical dispersion. To compensate for the possible increase of the dispersion one can modify the stiffness and the mass matrices A and M using modified quadrature rules as is done in [1].

In the m-adaptation approach, we consider a parameterized Mimetic Finite Difference (MFD) family of numerical schemes from which we select a member with the smallest numerical dispersion and anisotropy [2]. The parameters in the MFD family appear through the elemental mass and stiffness matrices M_E^{MFD} and A_E^{MFD}, respectively. The elemental mass matrix M_E^{MFD} on a square element E depends on two parameters m_1, m_2 while the elemental stiffness matrix A_E depends on one parameter ζ.

The MFD family parameterized by (m_1, m_2, ζ) contains a large number of known methods as special cases–for example, standard Finite Difference (FD), rotated FD, weighted combination of standard and rotated FD, Finite Element (FE) with lumped mass matrix, and the modified quadrature method of Guddati and Yue [1]. Moreover, compared with the last method, the MFD family is richer–it contains one extra parameter.

For the acoustic wave equation in 2D the optimal parameters (m_1, m_2, ζ) can be selected based on the von-Neumann analysis. One

Fig. 2. Displacement as a function of the distance from the origin at time T = 0.9 obtained using the modified quadrature method (top) and the m-adaptation method (bottom) for a Gaussian initial displacement data.
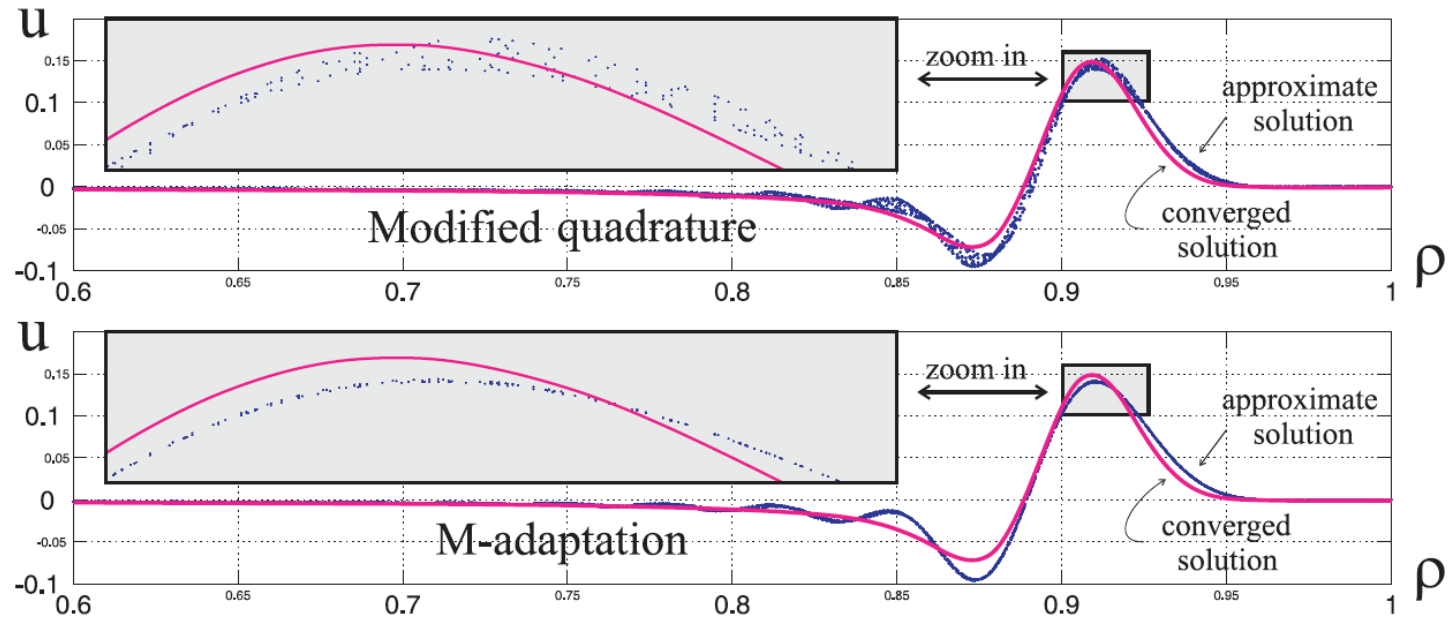
obtains a local dispersion equation relating the numerical velocity of the wave c_h with its wave number κ, mesh size h, and the parameters (m_1, m_2, ζ). Expending the error between the physical and numerical velocities of the wave, $c - c_h$, in powers of wave resolution number, κh, we select the parameters (m_1, m_2, ζ) to eliminate the error at the leading powers of κh. As a result of m-adaptation, the numerical velocity c_h is accurate to the fourth order in dispersion (as in [2]) and to the sixth order in anisotropy (versus fourth order in[2] see Fig. 1 and 2).

In the future we plan to develop the m-adaptation technique for higher order schemes on general meshes and for elastic wave equations. The potential of m-adaptation is high because with increased order of the scheme and/or number of vertices in the element, the number of free parameters grows quadratically. This may lead to a dramatic improvement in the dispersion and anisotropy of the optimal scheme.

[1] Guddati, M.N. and B. Yue, *Comput Meth Appl Mech Eng* **193**, 275287 (2004).
[2] Gyrya, V. and K. Lipnikov, "M-adaptation Method for Solving Acoustic Wave Equation on Rectangular Meshes," *J Acoust Soc Am*, submitted (2012).

Nonlinear Cascades in Rotating Stratified Boussinesq Flows

Hussein Aluie, CNLS/T-5; Susan Kurien, T-5

We employ state-of-the-art numerical simulations hand-in-hand with a very general mathematical approach to analyze nonlinear scale interactions in geophysical flows. We use high-resolution, high-Reynolds number direct numerical simulations of rotating stratified Boussinesq flows carried out on the 0.6-petaflops, 164-K compute cores of the IBM Blue Gene/P at ANL. We study the transfer of quadratic invariants, energy and potential enstrophy between different scales in three different regimes of stratification and rotation. Our results include the first reported measurement of potential enstrophy flux across scales. They also show joint cascades of energy and potential enstrophy from large to small scales in marked departure from 2D turbulence, which is often used as a model of geophysical flows.

Sea-water density in the oceans is ≈ 1 g/cm^3 with relatively small variations as a function of depth (of order 10^{-2} g/cm^3 or less), and even smaller fluctuations due to fluid motion (of order ≈ 4 km or less) [1]. This makes the Boussinesq equations a very accurate description of the flow dynamics in Earth's oceans. Yet the equations are prohibitively expensive to simulate–which is why most of today's climate models of the ocean rely on further simplifications. Such reduced models are a very good description for the horizontal quasi-2D fluid motions at scales much larger than the ocean depth of ≈ 4 km, but do not capture the 3D turbulent fluid motions at smaller length scales. Furthermore, their predictive power fails on long time scales, over which the effect of small scales on the global flow dynamics cannot be neglected and is, in fact, an essential factor in phenomena such as the thermohaline circulation. Such shortcomings of existing models demonstrate the need for a better physical understanding of the nonlinear coupling between the motions at different length scales, and in various regimes of rotation and stratification in turbulent Boussinesq flows. The progress we made on this front [2] relied on two key tools: state-of-the-art numerical simulations and a powerful mathematical technique to probe the data.
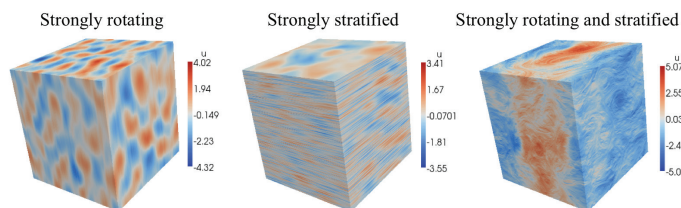
Thanks to a DOE "Innovative and Novel Computational Impact on Theory and Experiment" award (PI: S. Kurien), we carried out a series of computationally intensive Boussinesq simulations (Fig. 1) on the IBM Blue Gene/P at the ANL Leadership Computing Facility. Details of the simulations are described in [3].

We also used a novel and very general mathematical approach, introduced by [4], to study fluid flows. It was further refined and extended to magnetized [5], geophysical [2], and compressible [6] flows. The method itself is simple. For any field $\mathbf{a(x)}$, a "coarse-grained" or (low-pass) filtered field, which contains modes at length-scales $> l$, is defined as:

$$\bar{a}_l(x) = \int dr\, G_l(r)\, a(x+r)$$

where $G_l(r)$ is a normalized convolution kernel. The coarse-graining operation may be interpreted as a local space average. Coarse-grained dynamical equations can then be written to describe the large-scale flow at every point x in the domain, at every instant in time t, and at scales $> l$, for arbitrary l (Fig. 2). The approach, therefore, allows for the simultaneous resolution of dynamics both in scale and in space. Moreover, coarse-grained equations describe a range of large scales whose dynamics is coupled to the small-scale through so-called sub-scale terms. These terms depend inherently on the unresolved dynamics that has been filtered out. Traditional modeling efforts focus on devising closures for such terms that are plausible but whose regime of applicability and validity is inevitably unknown. A key feature of our formalism, which distinguishes it from such modeling efforts, is in developing a physical understanding of the sub-scale terms and estimating their contributions as a function of the resolution scale L through exact mathematical analysis and direct numerical simulations. The approach allows for studying the nature of coupling that exists between different scales and for unraveling certain scale-invariant or universal features in the dynamics. Furthermore, while traditional modeling efforts seek to formulate faithful models to describe the

Fig. 1. Boussinesq flow visualizations in three distinct regimes of rotation and stratification.

Strongly rotating | Strongly stratified | Strongly rotating and stratified
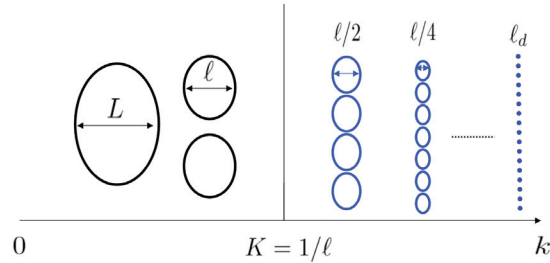
Fig. 2. A schematic depiction of the coarse-graining approach. Length scales in the system are partitioned into large, $\geq l$, and small, $< l$. Wavenumber, k, is inversely proportional to length scale; $k\sim l^{-1}$. The largest scale, that of the system's size, is L. Below scale l_d the dynamics is linear and modes are uncoupled. Length l represents the smallest scale that is resolved after coarse-graining. Scales $< l$ (in blue) are averaged out.

largest scales possible, typically of order L, our interest is to probe *all scales in the system*, including limits of small l << L. For example, in large-scale ocean simulations, the dynamical resolution is often variable in space due to adaptive mesh techniques and it is, therefore, imperative that sub-scale models used in these codes accurately reflect the latent physics as a function of resolution. Fundamental studies using the coarse-graining approach have a direct bearing on such issues and their findings may be naturally and easily translated to those modeling efforts in climate research.

We showed in [2] that, contrary to common belief, turbulent flows under strong rotation and/or stratification exhibit salient departures from 2D turbulence. In particular, an important finding of our work showed the existence of a joint cascade of energy and potential enstrophy, another dynamical invariant, from large to small scales (Fig. 3). Furthermore, these joint cascades exhibited pronounced but similar degrees of anisotropy, which are absent from the treatment of turbulent dissipation in current ocean models. The right three panels in Fig. 3 are the first measurements of potential enstrophy flux in rotating stratified Boussinesq flows. They can be regarded as the first empirical confirmation of analytical results in [7] which derived an exact law for potential enstrophy flux in physical space as a function of scale.
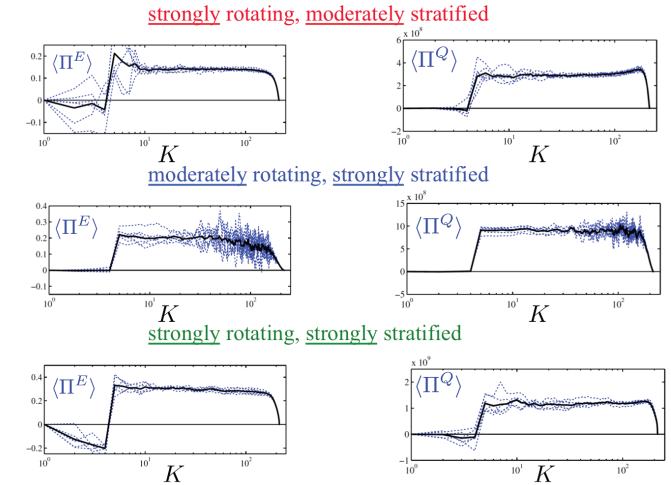
Fig. 3. Plots showing joint cascades of energy and potential enstrophy from large to small scales in all three regimes of rotation and stratification.

[1] Salmon, R., *Lectures on Geophysical Fluid Dynamics*, Oxford University Press, New York, New York (1998).
[2] Aluie, H., and S. Kurien, *Europhys Lett* **96**, 44006 (2011).
[3] Kurien, S., *J Fluid Mech*, submitted; arXiv:1005.5366 (2012).
[4] Eyink, G.L., *Physica D* **207**, 91 (2005).
[5] Aluie, H., and G.L. Eyink, *Phys Rev Lett* **104**, 081101 (2010).
[6] Aluie, H., *Phys Rev Lett* **106**, 174502 (2011).
[7] Kurien, S. et al., *J Fluid Mech* **555**, 131 (2006).

Compressible Turbulence: The Cascade and Nonlinear Scale Interactions

Hussein Aluie, CNLS/T-5; Shengtai Li, T-5;
Hui Li, T-2

We combined rigorous mathematics and very high resolution simulations to analyze nonlinear scale interactions of physical processes in compressible turbulent flows. We utilized and refined a novel coarse-graining approach, rooted in a commonly used technique in the subject of partial differential equations (PDE), to allow for the analysis of scale-coupling in highly compressible flows, and subsequently investigated the transfer of kinetic energy between different scales. We also carried out state-of-the-art high-resolution simulations of compressible turbulence on LANL's HPC clusters. By applying our mathematical tools to numerical data, we obtained the first empirical evidence that mean kinetic energy cascades conservatively beyond a transitional "conversion" scale range, despite not being an invariant of the compressible flow dynamics.

Turbulence is a phenomenon that pervades most liquid, gas, and plasma flows in engineering and nature. It often involves a huge range of spatio-temporal scales that cannot be simulated directly with today's computational resources–a limitation which is unlikely to cease in the foreseeable future. It is therefore imperative that we achieve a solid fundamental understanding of such flows in order to formulate realistic models. While the traditional Richardson-Kolmogorov-Onsager picture is a successful theory of incompressible turbulence, most flows in nature and engineering are characterized by significant compressibility effects. These range from high-speed flight, to nuclear fusion power reactors, combustion, star formation in molecular clouds, and supernovae.

Compared to incompressible flows, our theoretical understanding of compressible turbulence is still underdeveloped. It is known that such flows involve important physical processes such as shock formation, sound waves, distinct dissipation mechanisms, and additional nonlinearities that can significantly influence the dynamics and statistics of compressible turbulence.

Using tools from the field of partial differential equations (PDE), we were able to make tangible progress in understanding some of these basic physical processes [1,2]. A primary goal of PDE analysis is the formulation of models that can faithfully capture the effect of small (fast) scales on large-scale (slow) dynamics without explicitly resolving all degrees of freedom in the system. A novel and very general approach, inspired by methods from PDE analysis, was introduced by [3] to study nonlinear scale interactions in flow fields. It was further refined and extended to magnetohydrodynamic [4] and geophysical [5] flows. Guided by physical and mathematical considerations, we recently showed [1] how a Favre or density-weighted decomposition can be employed to extend the coarse-graining approach to highly compressible turbulence. A Favre filtered field is weighted by density as:

$$\tilde{a}_l(x) = \int dr\, G_l(r)\rho(x+r)a(x+r)\, / \int dr\, G_l(r)\rho(x+r)$$

where $G_l(r)$ is a normalized convolution kernel.

By employing this mathematical technique to analyze numerical data, we were able to gain new insights into the physics of compressible turbulence [2]. In inviscid compressible flows, only the sum of kinetic and internal energy is a global invariant. The idea of a cascade, a central notion in incompressible turbulence, is therefore without physical basis since kinetic energy is not conserved separately. Compressible flows allow for an exchange between kinetic and internal energy through two mechanisms: viscous dissipation and pressure dilatation. While the former process is localized to the smallest scales as in incompressible turbulence, the latter is a hallmark of compressibility and can a priori allow for an exchange at any scale through compression and rarefaction. By applying the coarse-graining technique to high-resolution 3D simulations, we obtained the first direct evidence that mean kinetic energy cascades conservatively beyond a transitional scale range, despite not being an invariant of the compressible flow dynamics [2]. The key quantity we measured is pressure dilatation co-spectrum (Fig. 1). The plot shows that it decays at a rate faster than k^1 as a function of wavenumber. This is sufficient to imply that mean pressure dilatation acts primarily at large scales and vanishes at small scales beyond the transitional "conversion" scale-range. At these
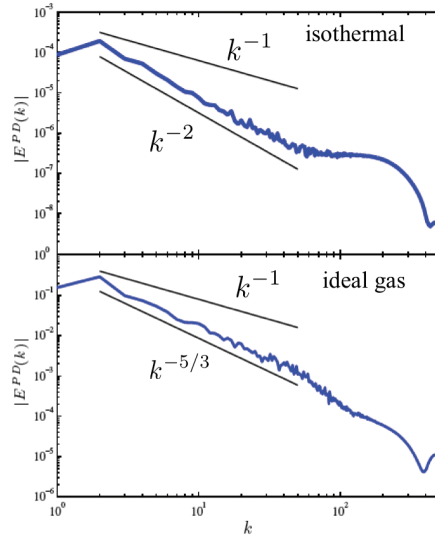
Fig. 1. Plots of pressure dilatation co-spectrum, showing a decay faster than k^{-1} as a function of wavenumber in isothermal (top) and ideal gas (bottom) turbulent flows. This is sufficient to imply that mean pressure dilatation acts primarily at large scales and vanishes at small enough scales, over which mean kinetic and internal energy budgets statistically decouple and mean kinetic energy undergoes a scale-local conservative cascade.
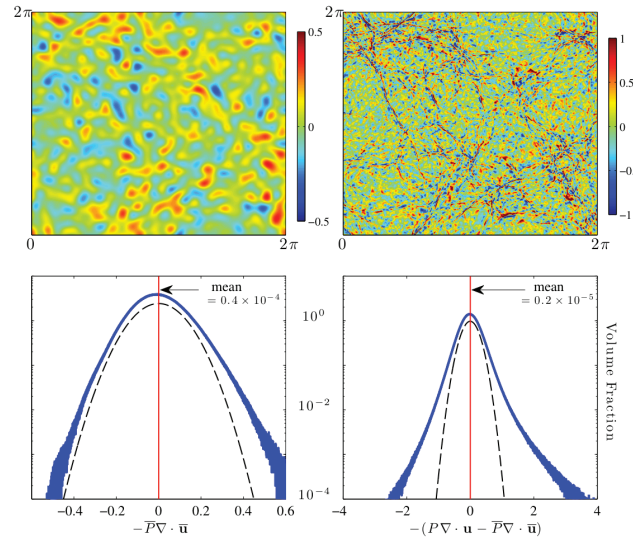
Fig. 2. Pointwise visualization of pressure dilatation in a 1024^2 cross-section of our 3D simulation at large (top-left) and small (top-right) scales. Lower two panels show the respective distribution of pressure dilatation values in the simulation domain. While pressure dilatation has large values at small scales in the vicinity of shocks, such small-scale contributions vanish when averaging over the flow domain due to cancellations between compression and rarefaction regions.

smaller scales, mean kinetic and internal energy budgets statistically decouple, giving rise to an inertial range over which mean kinetic energy undergoes a scale-local conservative cascade.

Our work is based on a rigorous mathematical analysis of the dynamics supported by high-resolution numerical simulations. The results and the physical picture we are advancing might seem counterintuitive at first. After all, a hallmark of compressible turbulence is the formation of shocks and the generation of sound waves. Such phenomena involve compression and rarefaction at all scales and are not restricted to large scales. However, our results concern global pressure dilatation, $-\langle P\nabla \cdot u \rangle$, and not the pointwise quantity. We showed (Fig. 2) that while pressure dilatation has large values at small scales in the vicinity of shocks, such small-scale contributions vanish when averaging over the flow domain due to cancellations between compression and rarefaction regions. Our findings should have significant implications on devising reduced models of compressible turbulence as well as providing physical insight into this rich problem.

[1] Aluie, H., *Phys Rev Lett* **106**, 174502 (2011).
[2] Aluie, H. et al., *Astrophys J Lett*, submitted (2012).
[3] Eyink, G.L., *Physica D* **207**, 91 (2005).
[4] Aluie, H., and G.L. Eyink, *Phys Rev Lett* **104**, 081101 (2010).
[5] Aluie, H., and S. Kurien, *Europhys Lett* **96**, 44006 (2011).

Direct Numerical Simulations of Rayleigh-Taylor Instability with 0.04 ≤ A ≤ 0.9

Daniel Livescu, Tie Wei, CCS-2

We have concluded our high-resolution, high-Reynolds-number Direct Numerical Simulations (DNS) of the Rayleigh-Taylor instability (RTI) on the 0.5-petaflop/s, 150-k compute cores BG/L Dawn supercomputer at LLNL. This represents a large set of simulations that include a suite of runs with Atwood numbers ranging from 0.04 to 0.9 and a multitude of initial conditions on a grid size of 1024^2 by 4608, as well as a high resolution run of grid size 4096^2 by 4032 and Atwood number of 0.75. After the layer width had developed substantially, additional branched simulations were run under reverse gravity and zero gravity conditions. The simulations provide an extensive database for the study of RTI turbulence and its dependence on the initial conditions, including mixing layer growth rates, turbulence and mixing asymmetries, self-similar behavior, and spectral characteristics. Individual terms of the moments transport equations have been recorded to develop and validate turbulence closure models. Here, we showcase the preliminary analysis of the data in connection with the growth of the layer and the turbulence problem.

Fig. 1. Rayleigh-Taylor bubble growth rate α from several simulations with A=0.04 and different initial perturbation spectra, shown in the legend. DNS with appropriate perturbation spectrum can reproduce the higher growth rates seen in the experiments.
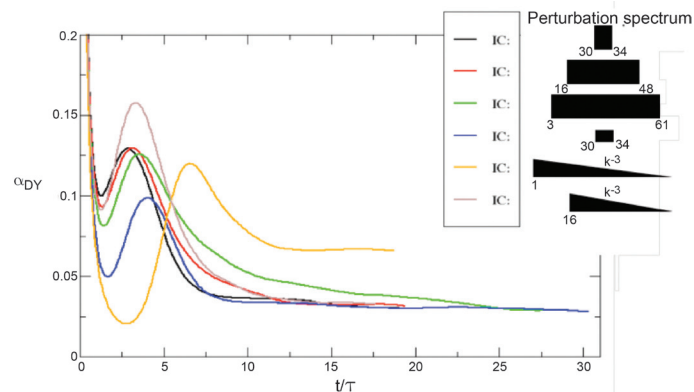
Molecular mixing in response to stirring by turbulence is an important process in many practical applications. When the microscopic densities of the fluids participating in the mixing are very different, these flows have been referred to as variable density (VD) flows, in contrast to the Boussinesq approximation in which the densities are commensurate [1]. Fundamental turbulence studies as well as specific engineering models for such VD flows are scarce [2,3]. Many of these flows are driven by acceleration (e.g., gravity in geophysical and astrophysical flows) which, because the density is not uniform, leads to large differential fluid accelerations. If the acceleration is constant and the fluid configuration is unstable (i.e., the density gradient points opposite to the acceleration), a fluid instability is generated in which small perturbations of the initial interface between the two fluids grow, interact nonlinearly, and lead to turbulence. This instability is known as the Rayleigh-Taylor instability (RTI) and is of fundamental importance in a multitude of applications, from fluidized beds, oceans and atmosphere, to inertial, magnetic, or gravitational confinement fusion, and to astrophysics.

Although this instability has been subjected to intense research over the last 50 years (e.g., over 100 RTI-related papers are published every year in peer-reviewed journals), early numerical studies have been restricted to coarse mesh calculations. On the other hand, it is notoriously difficult, experimentally, to accurately characterize the initial conditions and provide the detailed measurements needed for turbulence model development and validation. Thus, a large number of open questions remain unanswered about this instability and even first-order global quantities, such as layer growth, are not completely understood and still give rise to intense debate [4,5]. Nevertheless, today's petascale computers allow fully resolved simulations of RTI at parameter ranges comparable to those attained in laboratory experiments, but providing, in carefully controlled initial and boundary condition studies, much more information than the actual experiments [2]. These extremely high-resolution simulations are enabling a look at the physics of turbulence and testing turbulence models in unprecedented detail, hopefully contributing to a significant advance in our predicting capabilities.

To examine the turbulence and mixing properties, provide data for model development and testing, and explore the influence of the initial conditions, we have performed fully resolved, very-high-resolution simulations of Rayleigh-Taylor instability with the CFDNS code [6]. These simulations are the largest fully resolved simulations of the RTI to date and cover the range of Atwood numbers A=0.04 to 0.9 in order to study small departures from the Boussinesq approximation
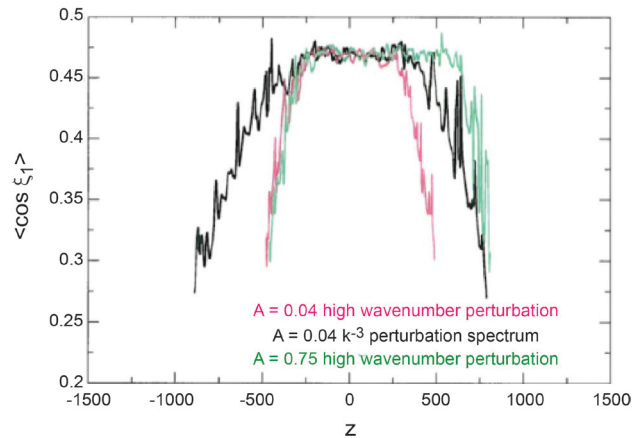
Fig. 2. The variation across the layer of the mean cosine of the angle between the vorticity vector and the most extensional eigenvector of the strain rate tensor. The interior turbulent regions do not depend on the initial conditions for a given A and the vorticity alignment is similar to other canonical turbulent flow. The edge regions have clear non-turbulent behavior and their extent depends on the initial conditions.

as well as large Atwood number effects, which are far less understood. Thus, most previous studies address the low-to-moderate A and no DNS have been reported, prior to the present set, for A>0.5. At high A, the velocity field is no longer solenoidal even when the two fluids are incompressible. The development of the instability and the mixing itself are fundamentally different at high and low A [1,5]. After the layer width had developed substantially, additional branched simulations were run under reverse gravity and zero gravity conditions. This "gravity reversal" occurs in practical situations (e.g., inertial confinement fusion [ICF] or pulsating stars), however there are no fully resolved simulations to date in this configuration.

Our preliminary analysis of the data has shed new light on the long-standing open question regarding the discrepancy between the numerically and experimentally computed mixing layer growth rates [7]. Thus, the results point to two possible explanations for the higher growth rates measured in various experiments (either one or both explanations can apply): 1) experimental set-up too small and thus the measurements represent only the early time behavior, and/or 2) layer growth affected by the lateral walls due to large wavelengths present in the initial perturbation spectrum. These hypotheses are also supported by two recent experimental studies with carefully controlled initial conditions [8,9]. An important finding resulting from our extensive dataset is that we were able to identify classes of initial conditions which can lead to faster-mixing layer growths, comparable to the experimental values.

In general, higher self-similar growth rates are obtained if the initial perturbation spectrum contains significant low-wave-number content, such as a spectrum of the type k^{-3} starting at $k=1$. In this case, the instability is laterally confined (the horizontal scales cannot grow larger) and the growth mechanism is different than mergers of bubbles

and spikes, as previously thought to characterize RTI layer growth. In fact, the behavior near the edges of the dominant bubble/spike is very similar to what we discovered in a previous study occurs in single-mode RTI at late times [10]. In this case, the vortexes generated on the sides of the bubble/spike self-propagate towards the edges of the layer and the induced velocity due to these vortexes helps the instability growth. This mechanism occurs for the other classes of initial conditions as well, however the vortical motions are influenced by the vortices in the nearby bubbles/spikes and the induced velocities are much weaker. Nevertheless, even for these cases, the merger growth mechanism is still less important for the overall growth. Further investigation of the edge regions has also shown, for the first time, that the mixing layer consists of a fully turbulent inner region and a turbulent/non-turbulent interface near the edges. The turbulent region becomes independent of the initial conditions for a given A, while the turbulent/non-turbulent interface retains the memory of the initial conditions and can grow at different rates for various classes of initial conditions. This shows that the mixing layer width measure in use today needs to be updated and establishes a minimum growth rate for the overall layer width, as that given by the inner turbulent region.

[1] Livescu, D. and J.R. Ristorcelli, *J Fluid Mech* **605,** 145 (2008).
[2] Livescu, D. "Numerical Simulations of Two-fluid Turbulent Mixing at Large Density Ratios and Applications to the Rayleigh-Taylor Instability," *Proc R Soc A,* to appear (2012).
[3] Schwarzkopf, J.D. et al., *J Turbul* **12,** N48 (2011).
[4] Livescu, D. et al., *J Turbul* **10,** N13 (2009).
[5] Livescu, D. et al., *Phys Scripta* doi:10.1088/0031-8949/2010/T142/014015 (2010).
[6] Livescu, D. et al., "CFDNS: A Computer Code for Direct Numerical Simulations of Fluid Turbulence," LANL Technical Report LA-CC-09-100 (2009).
[7] Livescu, D. et al., *J Phys Conf* **318,** 082007 (2011).
[8] Carles, P. et al., Phys Rev Lett 96 10450 (2006).
[9] Olson, D.H. and J.W. Jacobs, *Phys Fluid* **21,** 034103 (2009).
[10] Wei, T. and D. Livescu, "Initial Conditions and Reynolds Number Effects on Single-mode Rayleigh-Taylor Instability," *Phys Fluid*, submitted (2011).

Turbulent Flow Predictions for Grid-to-Rod Fretting in Pressurized Water Reactors

Mark A. Christon, Jozsef Bakosi, Nathan Barnett,
Marianne M. Francois, Robert B. Lowrie, CCS-2

Computational fluid dynamics simulations are described for the grid-to-rod fretting (GTRF) problem that occurs within the core of a pressurized water reactor. The simulations are used to predict the turbulent excitation forces on the reactor fuel rods. These forces vibrate the rods and may result in rod wear and nuclear fuel leakage. A description of the overall GTRF problem is given here, along with sample turbulent-flow-simulation results.

Within the core of a nuclear pressurized water reactor (PWR), water flow is used to cool the fuel rods and extract thermal energy. The grid-to-rod fretting (GTRF) problem in such reactors is a flow-induced vibration problem that results in wear and failure of the rods. GTRF wear is one of the leading causes of leaking nuclear fuel and costs power utilities millions of dollars in preventive measures. In order to understand the root causes of such fuel leaks, this research investigates the complex turbulent coolant flow around a fuel-rod bundle. Ultimately, our goal is to accurately predict the turbulent excitation forces on the fuel rods, along with the coupled structural response of the rods and their supports. To date, it has not been possible to completely characterize the flow-induced fluid-structure interaction (FSI) problem for GTRF. Indeed, given the incompressible nature of the coolant, the relatively high Reynolds number, and the flexible character of the fuel rods and spacers, the FSI problem at the reactor core scale is daunting.

Our investigation centers on large-scale computational fluid dynamics simulations, using the Hydra-TH code, which is being developed by our team at LANL. Hydra-TH has the capability of computing high Reynolds-number turbulent flows over complex geometrical objects. There are tens of thousands of fuel rods in a PWR, and to compute the flow over all of them simultaneously, with the fidelity required for GTRF, is not feasible on today's computers. Instead, this study concentrates on a representative 3×3 rod bundle, for which a sample computational mesh is shown in Figs. 1 and 2. The 3×3 geometry was extracted from a 17×17 fuel assembly that is found in a typical

Fig. 1. Surface mesh for 3×3 rod bundle. Only the center rod is shown, along with the grid spacer geometry.

PWR. There is a large degree of symmetry in the fuel assembly, which makes this geometrical simplification a reasonable approximation. The coolant flow generally moves axially down the outside of the fuel rods. A "grid spacer" (rod support structure) is also shown, which contains mixing vanes that stir the flow in order to enhance heat transfer from the rod to the coolant flow. This stirring, along with the high Reynolds number of the flow, results in a complex turbulent flow that is difficult to predict accurately.

Figures 3 and 4 show sample results from Hydra-TH, using the detached-eddy simulation (DES) turbulence model. In Fig. 3, vortex structures generated by the mixing vanes are shown as they transport down the rod. At an axial location downstream of the grid spacer, a cross-cut of the flow is shown in Fig. 4, where the details of the vortex structures are readily apparent. Although these vortex structures enhance the needed cooling, they also result in pressure fluctuations on the surface of the fuel rod, which in turn causes the fuel rods to vibrate. These vibrations cause the supports for the fuel rod to wear into the protective cladding on the surface of the rod, which, given enough time, may expose the internal nuclear fuel and result in a fuel leak. Understanding these interactions is critical to mitigating issues with GTRF. (For more examples and details on these calculations, see [1].)

Future work on GTRF will focus on coupling in the structural response of the fuel rods, along with coupling with wear models being developed by collaborators in the Consortium for Advanced Simulation of Light Water Reactors (CASL) project, which include research staff from MST and T divisions at LANL. The largest mesh run to date for the 3×3 problem is approximately 12 million elements. In order to adequately resolve the turbulent flow features, we believe a mesh of 100 million to 1 billion elements will be required. To adequately resolve the temporal
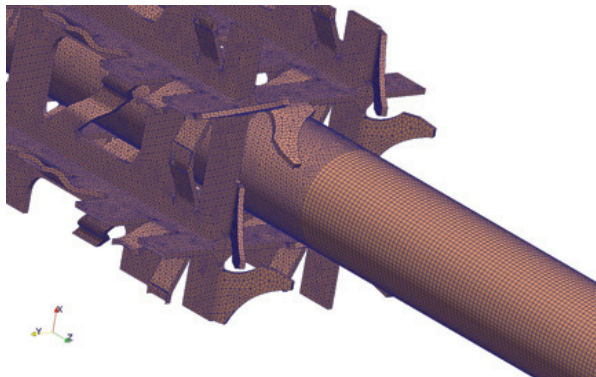
fluctuations, necessary to predict the vibration of forces on the rod, a large eddy simulation of the turbulent flow will be performed. The time-dependent nature of the flow, complex geometry, and high Reynolds number drives the requirement for large meshes. Generation of these meshes is currently underway. To address flows with boiling, multiphase flow models are also being developed for implementation in the flexible Hydra-TH software toolkit.

CASL, a DOE Innovation Hub led by ORNL, supports this research. LANL is full partner in the CASL Hub, along with SNL and INL, and other university and industrial partners (http://www.casl.gov/). The computations for this study were performed using LANL's Institutional Computing program.
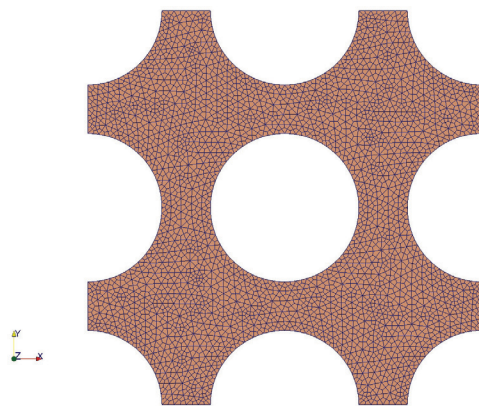
Fig. 2. Axial cross-section of computational mesh of coolant channels for a 3×3 rod bundle, downstream of the grid spacer.
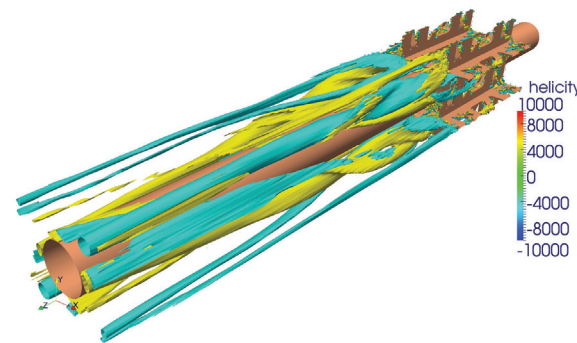
Fig. 3. Instantaneous isosurfaces of helicity for the 3×3 rod bundle problem.
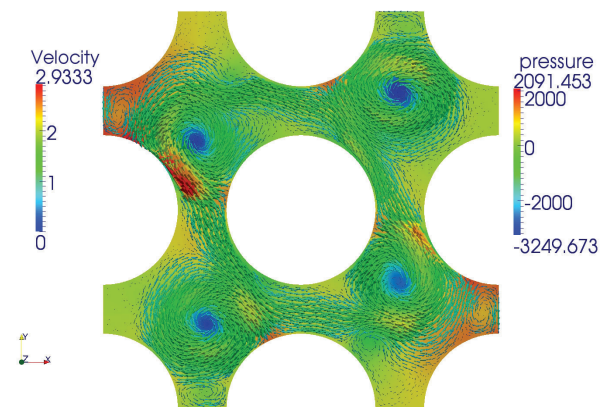
Fig. 4. In-plane velocity vectors, shading colored by pressure value, at an axial location just downstream of grid spacer, corresponding to results in Fig. 3. Mixing vanes at the grid spacer.

[1] Christon, M.A. et al., "Initial Assessment of Hydra-TH Code on GTRF Problems," LANL Report LA-UR-11-07034 and CASL milestone L2:THM.P4.01 (2011).

Uncertainty Quantification for Carbon-capture Simulation

Leslie M. Moore, K. Sham Bhat,
Joanne R. Wendelberger, CCS-6; David Mebane,
National Energy Technology Laboratory

The Carbon Capture Simulation Initiative (CCSI) is developing tools to accelerate identification of reliable and affordable processes for carbon capture from coal-fired power plants using simulation. The effort includes implementation of tools for uncertainty quantification (UQ), methodology critical to simulation-based analysis of complex processes, including economic impact of incorporation of carbon capture systems in current and future commercial operations. UQ tools include input sensitivity analysis, calibration of input parameters, construction of surrogate models, and propagation of uncertainty. Here UQ use is illustrated with study of a solid sorbent process for carbon capture, using preliminary models and thermogravimetric analysis (TGA) data from the National Energy Technology Laboratory (NETL).

Carbon capture and storage (CCS) technology promises to convert fossil fuels into a reliable low carbon energy supply while remaining affordable for consumers. The Carbon Capture Simulation Initiative (CCSI) will develop a simulation toolset supporting the needs of industry to evaluate new carbon-capture technologies, scalable to commercial level with reduced physical testing, including uncertainty quantification (UQ) tools to aid in interpreting simulation results. Solid sorbent modeling was chosen to demonstrate development and validation of computer simulation approaches.

UQ includes analytical tools used to understand variable processes at all levels of a system and to focus resources on uncertainty with large impact on full-scale system performance. CCS technology and its implementation include many processes with complex impacts on energy economics and the environment, so UQ is critical. Solid sorbent adsorption of carbon dioxide (CO_2) is a small-scale process that feeds into other CCS efforts such as basic chemistry, particle and device scale models, and plant process models (Fig. 1). UQ used in solid sorbent adsorption study supports understanding simulation performance, such as identifying parameters that induce more variation than others (sensitivity), estimating values for unknown parameters consistent with physical measurements, and quantifying parameter uncertainty from measurement error to model error sources.

Computer models of varying complexity simulate adsorption of CO_2 by a sorbent under well-known chemical and microstructural assumptions, with complexity dependent on extent of incorporation of

chemical, microstructure, and transport phenomenon. The sorbent is a mesoporous silica backbone embedded with the amine polymer, polyethyleneimine (PEI). CO_2 is adsorbed through two chemical reactions: 1) a CO_2 molecule binds with a PEI site forming a zwitterion, and 2) a zwitterion binds with another empty PEI site forming carbamate. Diffusion of gaseous CO_2 in the sorbent is explained by a microstructural model divided into three length scales with different diffusion types. Infinitely fast gas diffusion and Knudsen diffusion occur at the large and middle length scales, while bulk-phase diffusion, where zwitterions hop between PEI sites over an energy barrier, occurs at the smallest length scale. Two computer simulation models of solid sorbent technology are explored: an ideal equilibrium model with five parameters, and a more complex second generation model implementing basic transport, kinetic, and ideal thermodynamics in a dry environment, with twelve parameters. The parameters, not measurable directly, are estimated by calibrating to thermogravimetric (TGA) experiments for a dry sorbent from the National Energy Technology Laboratory (NETL). TGA measurements are sorbent weight versus a temperature profile changing over time at specific CO_2 composition (Fig. 2 at 10 %). Many sources of uncertainty exist in this framework: data error, inaccurate modeling assumptions, boundary conditions, extrapolation, and model scaling. UQ demonstration includes parameter sensitivity assessment for a second generation model and parameter calibration illustrated for an ideal equilibrium model.

The second generation model simulates a TGA curve for a set of input parameters, an assumed temperature profile and CO_2 pressure. UQ tools in model development provide early parameter studies and identify unexpected performance of a simulation code. An initial parameter study

Fig. 1. Scales for process engineering, showing the need for a scale-up approach to simulation.
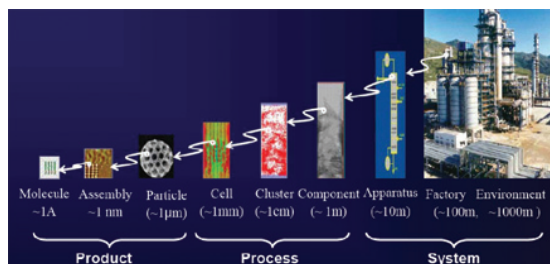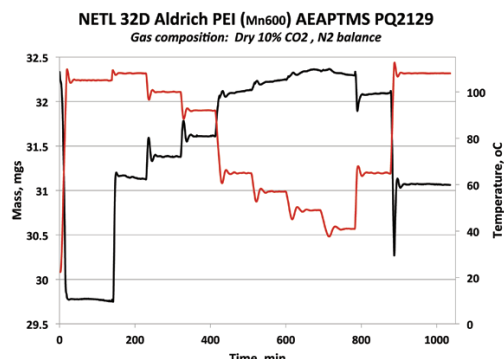
Fig. 2. Thermogravimetric analysis (TGA) response data (black) of sorbent weight measured every second, coincident with temperature profile. Temperature profile (red) includes seven temperatures and eight transitions over time, and same profile was used in model runs.

conducted with the second generation model used an ensemble of 128 sets of the twelve parameters, selected to achieve balanced representation of multiple factors at high and low levels with an orthogonal-array based Latin Hypercube Sample (LHS), a strategy in common use in computer experiments [1-3]. Sensitivity analysis was based on identifying trends in a simple scalar calculated from the TGA curves: mean weight fraction. Nearly half of the ensemble runs either did not complete or achieved very low peak weight fractions indicating model performance issues, and results were less informative of parameter values consistent with observed TGA data. Parameter sensitivity evaluation focused attention on two parameters with apparent main effects, affirming implementation of a model of sorbent kinetics controlled by number and mobility of zwitterions.

A simplified ideal equilibrium model calculates equilibrium weight for fixed temperature and five input parameters based solely on reactions of the sorbent, ignoring kinetic effects. Parameter estimation (calibration) determines the "true" input parameter set, that for which the model "best" matches observations, while accounting for model discrepancy, observation error, and any bias. Observed equilibrium weights were obtained from the seven temperature plateaus in the experimental data. A Bayesian approach [4] for parameter estimation provides a posterior probability distribution of the "true" input parameters inferred from model output, physical observations and specified prior parameter distributions. MCMC simulation produced sample-based posterior distributions. Numerically integrating over other parameters, a sample-based posterior distribution of the "true" input parameters is obtained with illustration as pairwise bivariate posterior distributions of the parameters in Fig. 3a. Predictions were obtained for each posterior sample, yielding a predictive distribution of sorbent weights at particular temperatures and pressure, with mean and 95% credible regions displayed graphically (Fig. 3b) with observed equilibrium weights in red showing model and data inadequacies.

The ideal equilibrium model is a substantive simplification of the current second generation model, itself a middle fidelity model incorporating
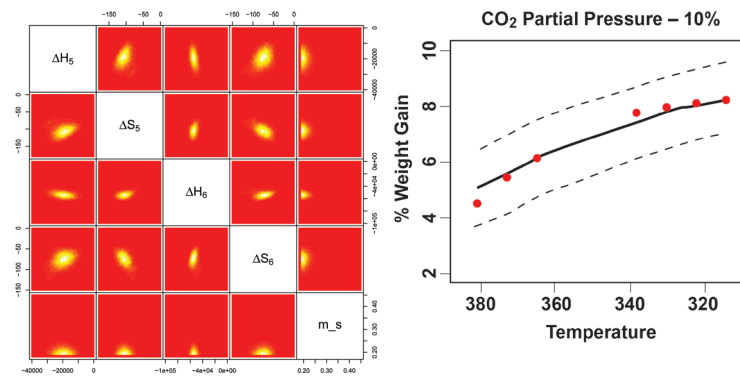
Fig. 3. a) Bivariate posterior probability densities illustrate "most likely" input parameter values. b) Simulation based prediction and bounds (95%) for 10% partial pressure, physical observations in red.

a transport model with ideal thermodynamics. Higher fidelity, more relevant models require non-ideal thermodynamics and a hydrous environment. Eventually, a solid sorbent process simulator with parameter values consistent with experimental data will be validated for use in higher-level simulations of CCS technology. UQ issues for CCS technology simulation dependent on a solid sorbent process simulator will include 1) whether the fidelity of the solid sorbent model has an important impact on CCS technology, and 2) how to quantify the impact. UQ will be used again to address such questions.

[1] Chen, V.C.P. et al., *IIE Transactions* **38,** 273 (2006).
[2] Santner, T.J. et al., *The Design and Analysis of Computer Experiments,* Springer, New York (2003).
[3] Williams, B. et al., *Bayesian Analysis* **1,** 765 (2006).
[4] Kennedy, M., and A. O'Hagan, *J Roy Stat Soc B Stat Meth* **68,** 425 (2001).

A Moment-based, Scale-bridging Algorithm for Neutral Particle Transport Problems

Hyeongkae Park, Rick M. Rauenzahn, Dana A. Knoll,
Christopher K. Newman, T-3; Allan B. Wollaber,
Jeffery D. Densmore, CCS-2

Accurate modeling of neutral particle behavior is of great importance in many science and engineering applications. Due to strong interactions between particles and surrounding media, many numerical algorithms suffer from slow convergence in optically thick regions. We are developing a moment-based, scale-bridging algorithm that can mitigate slow convergence of transport algorithms. Our algorithm utilizes the low-order (LO) continuum equations that are consistently derived from the high-order (HO) transport equation. Due to discrete consistency, the LO system can be used not only for accelerating the HO solution, but also coupling to other physics.
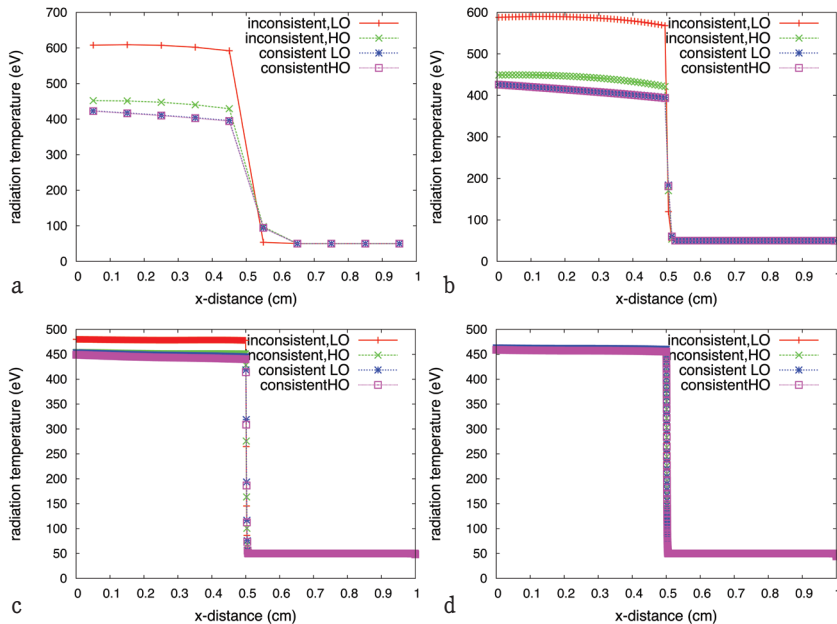
The Boltzmann transport equation describes the statistical behavior of particles that are interacting with surrounding media. The transport equation is used to simulate neutral particle physics in many physical systems–for example, combustion systems, nuclear power reactors, and internal confinement fusion. Unfortunately, many iterative schemes, based on an explicit scattering and re-emission source iteration, suffer from slow convergence in an optically thick region, where particles undergo a large number of scattering and/or absorption re-emission events. In such regions, a continuum description of the transport equation with an approximate closure (e.g., Fick's law of diffusion) can adequately represent the particle behaviors. On the other hand, diffusion and other approximated closure equations become invalid in an optically thin region due to the presence of significant transport effects. For computational efficiency

it is natural to seek a (geometric) hybrid transport-diffusion algorithm [1,2].

As an alternative to this geometric hybrid algorithm, we are developing a new, efficient approach to solve neutral particle transport problems. In our algorithm, the transport high order (HO) and continuum low order (LO) systems lie on the same computational domain, and can be viewed as a scale-bridging algorithm. Fine-scale physics, such as particle streaming, are treated by the HO system, while large-scale, isotropic physics are treated implicitly in the LO system. Our algorithm derives a closure relationship directly from the discrete HO system and adds a *consistency* term that forces the LO truncation error to be identical to the truncation error of the HO system. Thus, the solution of the LO system is guaranteed to be consistent with the solution of the HO system upon convergence of the nonlinear system.

With increasing interest in high-fidelity, multiphysics simulations, our algorithm becomes even more attractive. Specifically, the LO system can be used to couple other physics (e.g., fluid flow, heat transfer, and chemical reactions) in a multiphysics simulation. By coupling with the LO system, we can reduce the problem dimension of the multiphysics system. Furthermore, the algorithm allows one to use not only a deterministic but also a stochastic (Monte Carlo) method for obtaining the solution of the HO system. We have applied our algorithm to many physical systems, but will focus on the following two here. The first application is time-dependent thermal radiative transfer (TRT) problems in the context of high-energy density physics [3]. The second application is neutron transport criticality [4], which is an important application in the nuclear power industry.

Fig.1. Comparison of the radiation temperature with and without the consistency term for different spatial mesh sizes: a) dx = 0.1 cm, b) dx = 0.01 cm, c) dx = 0.001 cm, d) dx = 0.0001 cm.
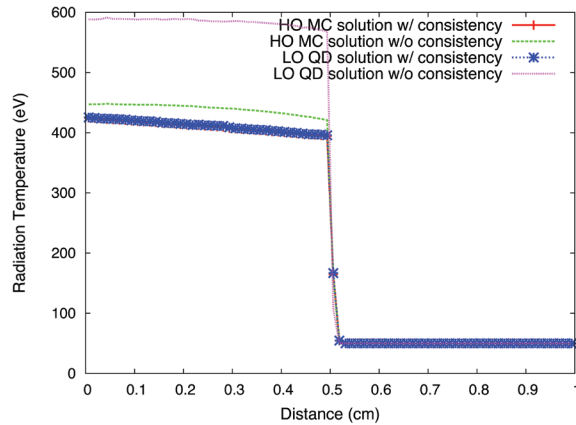
Fig. 2. Comparison of radiation temperature profile using Monte Carlo HO solver.
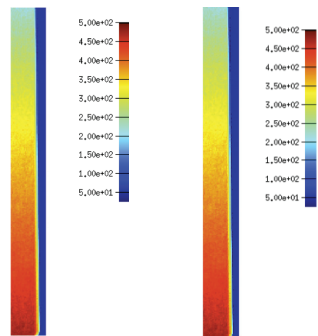
Fig. 3. Comparison of the radiation temperature in IMC HO solver: a) LO radiation temperature, b) HO radiation temperature.

Figure 1 shows the radiation temperature profiles for a 1D multimaterial TRT problem, solved with and without the consistent LO formulation. Without consistency, the LO radiation temperature exhibits an unphysical solution. This error is due to poor resolution of the solution gradient near the material interface. Because insufficient energy is transferred into the thick material, excessive energy is deposited in the thin material. By using the consistent LO formulation, the radiation flux at the material interface is properly evaluated, eliminating the unphysical solution. Although the LO and HO radiation temperatures approach each other without the consistency term under mesh refinement, mesh resolution on the order of a mean-free-path (not reasonable) is required to obtain a reasonable solution. Figure 2 shows the comparison as similar to Fig. 1, but the HO system is solved with a Monte Carlo method. With the Monte Carlo HO solver, the LO and HO solution are consistent. With the help of the LO problem, the particle tracking in the Monte Carlo simulation becomes simpler and faster, enabling a larger number of histories with the same CPU time. Figure 3 shows the radiation temperature profiles of a 2D multimaterial problem. The simulation was performed using a LANL large-scale multiphysics simulation code interfaced with an Implicit Monte Carlo (IMC) code. Both are production-level codes used regularly by a practitioner for realistic simulations. With the new algorithm, we observed CPU speed-ups of a factor of 8 to 20, compared to the standalone IMC code.

These same algorithms can be applied to neutron transport in nuclear power reactors. Figure 4 shows an eigenvalue convergence comparison between standard power iteration (PI) and our algorithm (NDA-NCA) in a 2D nuclear reactor criticality problem. The outer iteration corresponds to the number of HO solves, which is the major contributor to the CPU time. The NDA-NCA algorithm [4] reduces the outer iteration by a factor of 600 and reduces the CPU time by a factor of 400. Note that our algorithm not only accelerates scattering source via the scale-bridging

algorithm, but also utilizes an accelerated eigenvalue computation in the LO system, which is possible due to the consistent formulation between the HO and LO systems. The next step is multiphysics coupling with this scale-bridging algorithm [5].

Finally, this scale-bridging algorithm may map well to emerging parallel computing architectures. Specifically, the computationally expensive HO solution algorithm is suited for GPU computation, while we can capitalize on the state-of-art parallel computing software for solving the LO system across nodes. Moreover, a relatively small communication requirement between the HO and LO systems may ease the communication bottleneck of the heterogeneous parallel architecture.
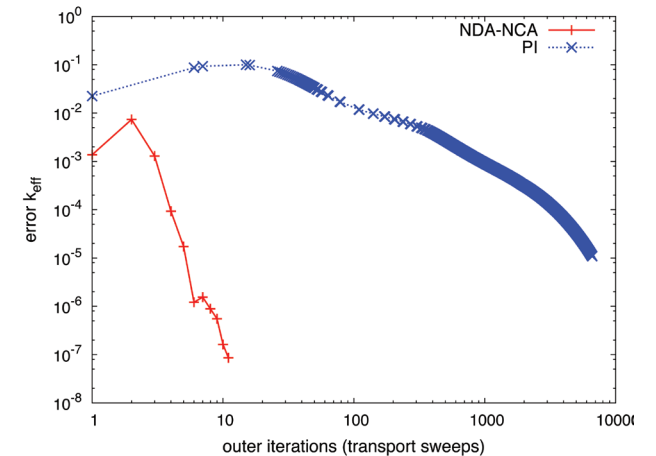
Fig. 4. Convergence comparison between NDA-NCA and PI for the LRA-BWR problem.

[1] Gentile, N., *J Comput Phys* **172**, 543-571 (2001).
[2] Densmore, J.D. et al., *J Comput Phys* **222**, 485-503 (2007).
[3] Park, H. et al., *Transport Theor Stat*, accepted for publication (2012).
[4] Park, H. et al., *Nucl Sci Eng,* accepted for publication (2011).
[5] Park, H. et al., *Nucl Sci Eng* **166,** 118-133 (2010).

Discrete Ordinate Calculation of the k-Eigenvalue of an IFBA Pin Using Unstructured Meshes in 2D

Massimiliano Rosa, James S. Warsa, Jae H. Chang, Randal S. Baker, CCS-2

Integral Fuel Burnable Absorber (IFBA) pins are used for the reactivity control of the fuel assemblies of light water reactors. From the standpoint of the feasibility of performing a pin-resolved, 3D neutron transport computation of a full reactor core, the presence of IFBA pins represents a challenging problem. The traditional approach of homogenizing cross-sections inside the cells of a Cartesian mesh can be used to capture the neutronics effects associated with the thin IFBA coating via a discrete ordinate (SN) transport computation of the k-eigenvalue of the pin. Unfortunately, the number of spatial mesh cells required to resolve the thin neutron absorbing layer in a single pin can become prohibitively high, even in only 2D. We show how the use of unstructured meshes in the SN computation drastically reduces (100-fold reduction) the number of cells required to obtain the same k-eff, while eliminating recourse to cross-section homogenization.

Integral Fuel Burnable Absorber (IFBA) pins are used for the reactivity control of the uranium dioxide (UO_2) fuel assemblies of light water reactors [1]. IFBA is applied as a thin coating of zirconium diboride (ZrB_2) on the outer surface of the UO_2 pellets [2]. Specifically, the cross-sectional geometry of the IFBA pin is sketched in Fig. 1. The extension of the coolant box is equal to the standard 1.26-cm pin pitch of a 17×17 fuel assembly in a pressurized water reactor. Resolving the IFBA coating layer, only 0.001-cm thick, makes the transport problem particularly challenging. The pin's material composition is described in Table 1 and is loosely based on the specifications given in [1].

Fig. 1. Geometry specifications for the IFBA pin.

Clad (0.476 cm)
Fuel (0.412 cm)
IFBA (0.413 cm)
He Gap (0.422 cm)

Table 1. Material composition of the IFBA pin.

Region	Density [g/cm³]	Material
Fuel	10.24	UO_2 4.2%: U^{235}: 4.2 wt%, U^{238}: 95.8 wt%
IFBA	1.69	ZrB_2
Gap	0.001	O^{16}
Clad	6.504	Zircaloy-2: Zr/Sn/Fe/Cr/N= 98.23/1.50/0.12/0.10/0.05 at%
Coolant	1.0	H_2O at 293.125 K and 101.325 KPa

The k-eigenvalue (k-eff) calculation via the homogenization approach [3] was carried out using the PARTISN parallel neutral particle SN transport code. Specifically, we performed a convergence study for the k-eff by considering increasingly refined Cartesian meshes over a quarter of the pin with reflective boundary conditions. We utilized multi-group neutron cross-sections from the "mendf6" Nuclear Data Interface (NDI) library, for the 30-energy-group structure "30_LANL" [4]. A Legendre order of three was selected for the scattering expansion. The angular quadrature was the S8 square Chebychev-Legendre set, kept constant to focus on the differences between spatial mesh refinement for Cartesian and unstructured meshes. The results of the "spatial" convergence study are reported in Table 2.

Table 2. Spatial convergence of "homogenized" k-eff.

Spatial Mesh	k-eff
16×16	0.63850091
32×32	0.64022925
64×64	0.64068658
128×128	0.64091920
256×256	0.64108409
512×512	0.64118889
1024×1024	0.64125049
1152×1152	0.64125726
1280×1280	0.64126246
1408×1408	0.64126647
1536×1536	0.64126974

The refinement was arrested when convergence was achieved within 0.00001, obtaining a k-eff of 0.64127. This result was obtained for the 1408×1408 mesh consuming a solution time of 8998 seconds on 256 processors of the Typhoon cluster. The total number of sweeps was 13042 (58 outers).

We created a set of PYTHON scripts for the generation of unstructured meshes that cover either the entire pin, or just a hemisphere or a quarter. Each material region is subdivided into a set of radial and azimuthal intervals, used to generate an unstructured mesh whose origin is coincident with the pin's center. The radial sector closest to the origin is subdivided into triangular finite
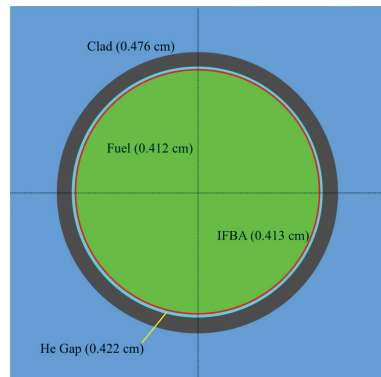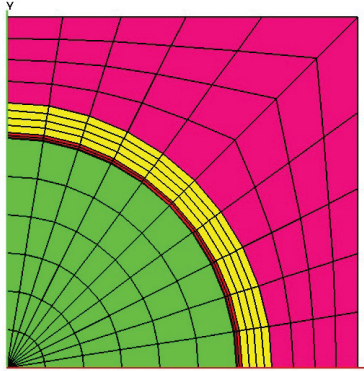
Fig. 2. A 17×10 unstructured mesh over a pin quarter.

elements. The remaining radial sectors are composed of quadrilaterals, to minimize the number of mesh cells. To fix ideas, a mesh example is given in Fig. 2.

The mesh in Fig. 2 is composed of 10 azimuthal intervals and 17 radial intervals: 6 in the fuel, 1 in the IFBA, 2 in the gap, 4 in the clad, and 4 in the coolant. Since no mesh cell contains more than one material, cross-section homogenization is not needed. The PYTHON scripts are also used to compute adjusted densities for the materials in the various regions of the meshed pin. This is done so that the mass of material in each meshed region is equal to the mass contained in the original curved volume for that region.

We used the CAPSAICIN parallel neutral particle SN code to compute the pin's k-eff over a fine unstructured mesh composed of 630 radial points: 412 in the fuel, 1 in the IFBA, 9 in the gap, 54 in the clad, and 154 in the coolant. The above choice results in a fine radial mesh where each radial interval is as thick as the radial extent of the IFBA coating. Using the same cross-section libraries and transport parameters as for the homogenization approach, we performed the azimuthal convergence study for the k-eff reported in Table 3.

Table 3. k-eff solution on a fine unstructured mesh.

Spatial mesh	k-eff
630×16	0.64130166
630×32	0.64128252
630×64	0.64127805
630×128	0.64127956
630×256	0.64128090

The converged k-eff is equal to 0.64128, and differs by 0.00001 from the "homogenized" value of 0.64127. The discrepancy stems from the fact that while CAPSAICIN exploits a bilinear discontinuous finite element method, PARTISN utilizes a linear discontinuous discretization.

We also investigated strategies for reducing the number of cells in the unstructured mesh computation. An encouraging result was obtained for a mesh composed of 228 radial intervals: 206 in the fuel, 2 in the IFBA, 4 in the gap, 8 in the clad, and 8 in the coolant. The k-eff was again 0.64128 and was already obtained for a 228×64 mesh, by switching from a linear to a logarithmically decreasing distribution of the radial points in the fuel, to appropriately resolve the transport effects near the fuel-IFBA interface. The computation took 1050 seconds on 240 processors. The total number of sweeps was 273, and a sweep includes all (30) energy groups. The above results were obtained exploiting CAPSAICIN's parallel decomposition in energy, using 30 energy processor groups. Finally, details on the non-linear formulation of the eigenvalue problem, and related solvers, used by the code can be found in [5]. The performance gain of the unstructured mesh strategy over the traditional homogenization approach is illustrated in Table 4.

Table 4. Performance of unstructured mesh approach.

	PARTISN	CAPSAICIN	Reduction Factor
Mesh cell count	1982464	14592	136
Solution time (seconds)	8998	1050	8.6
Sweep time (seconds)	0.69	0.13	5.3

The unstructured mesh calculation results in a 100-fold reduction in the number of mesh cells, consuming a solution time that is almost nine times lower than for the homogenization approach. We expect that the encouraging results obtained in 2D will favorably extend to higher dimensionality, and we will investigate the 3D case in our future research.

[1] Kozlowski, T., and T.J. Downar, https://engineering.purdue.edu/PARCS/MOX_Benchmark/.
[2] Franceschini, F., and B. Petrovic, *Ann Nucl Energ* **36,** 1201 (2009).
[3] Duderstadt, J.J., and L.J. Hamilton, *Nuclear Reactor Analysis,* Wiley, New York (1976).
[4] White, M.C., LANL Technical Report X-5MCW-03-50 (2003).
[5] Fichtl, E.D. et al., *Proc. M&C 2011*, Rio de Janeiro, Brazil (2011).

Direct Numerical Simulations of Isotropic and Post-Shock Turbulence Interacting with a Shock Wave

Jaiyoung Ryu, Daniel Livescu, CCS-2

We have conducted some of the largest Direct Numerical Simulations (DNS) to date of "isotropic turbulence" and "post-shock turbulence" passing through a normal shock wave. Such configurations are relevant in many practical applications and also represent unit problems for shock and re-shock phenomena in Richtmyer-Meshkov instability. We address for the first time the post-shock turbulence interacting with the shock wave. The inflow isotropic turbulence is generated using auxiliary forced compressible turbulence simulations with prescribed supersonic background speed. The inflow database for post-shock turbulence is collected from simulations of isotropic turbulence–shock wave interaction. Preliminary low-resolution results are presented here. We consider a shock wave with moderate strength (Mach = 1.2). Upstream of the shock wave, the microscale Reynolds number is 15, and the turbulence Mach numbers is 0.15 and 0.25. The behavior of the shocked and re-shocked turbulence is investigated by varying compressibility and Kolmogorov length scale of incoming turbulence. We present consistent results with previous studies on weakly compressible isotropic turbulence. Turbulent kinetic energy and spanwise vorticity fluctuations are amplified while passing through a shock wave, and the turbulence length scales decrease.

The interaction of shock waves with turbulence occurs in many interesting problems, including supernova explosions, inertial confinement fusion, and internal and external flow in hypersonic flight and propulsion. The coexistence of shock waves with some background turbulence is an almost unavoidable feature that creates complicating effects (anisotropy of the background turbulence, mean-flow gradients, oblique or bent shock wave, unsteady separation, etc.) so that a detailed understanding remains far from reach. For this reason, several studies [1–3] have focused on the interaction between isotropic turbulence and a normal shock wave as one of the basic phenomena involved in these flows. The purpose of this study is to produce flow statistics for turbulence modeling and to develop relevant theory for shocked and re-shocked turbulence by conducting accurate direct numerical simulations of corresponding problems.

In this study, all the relevant turbulence scales, as well as shock wave structures, are resolved as a solution of the Navier-Stokes equations without applying turbulence-model and shock-capturing methods using CFDNS code [4]. A plane steady shock wave satisfying Rankine-Hugoniot relations is specified as an initial condition in the middle of the domain. Compressible Navier-Stokes equations for a Newtonian fluid with perfect gas assumption are nondimensionalized by upstream density, temperature, and speed of sound. Prandtl number is taken as $Pr = 0.75$ and the ratio of specific heats is 1.4. It is important to

provide fully developed and realistic turbulence at the upstream of the shock wave. Incoming isotropic turbulence is generated using auxiliary forced compressible turbulence simulation with prescribed supersonic background velocity. In this manner we avoid the ambiguity and limitations of using Taylor's hypothesis, which is typically applied for temporally decaying isotropic turbulence to generate spatially decaying turbulence. We apply the linear turbulence forcing method [5–7] in triply periodic domain. After reaching a statistically stationary state, plane data in the middle of the domain perpendicular to the streamwise direction is collected and used as the inflow boundary condition for the isotropic turbulence interacting with shock wave simulation. Plane data behind the shock wave is collected and used as the inflow condition for another post-shock turbulence interaction with shock wave problem.

A range of isotropic turbulence simulations is conducted for various forcing wavenumbers and Kolmogorov length scales. Here we present preliminary results on the shocked and re-shocked turbulence. Figure 1 shows an instantaneous streamwise velocity field of isotropic turbulence and shock wave interaction. The velocity field is nondimensionalized with the speed of sound of upstream flow. The streamwise velocity fluctuates around a specified supersonic Mach number (1.2) in the upstream of the shock wave, and after being suppressed by the shock wave the velocity fluctuates around subsonic speed. Figure 2 compares the variation of Reynolds stress across the shock wave among different

turbulent Mach numbers and shocked and re-shocked turbulence. To investigate post-shock turbulence and shock wave interaction, plane data behind the shock wave is saved to match the turbulent Mach number of inflow isotropic turbulence. Red solid and black dot-dashed lines represent strong and weak post-shock turbulence, respectively, interacting with the shock wave. Blue dashed and violet dotted lines respectively represent strong and weak shocked isotropic turbulence. For both shocked and re-shocked turbulence, Reynolds stresses are less amplified for stronger turbulence. For a similar turbulence strength, isotropic turbulence is less amplified than anisotropic post-shock turbulence. This result is consistent with the synthetic anisotropic turbulence interacting with the shock wave in [8]. Further analysis will be conducted with the ongoing high-resolution simulations.

Fig. 2. Reynolds stress. The effect of changing turbulent Mach number and the difference between shocked and re-shocked turbulence. Red solid and black dot-dashed lines respectively represent strong and weak re-shocked turbulence. Blue dashed and violet dotted lines respectively represent strong and weak shocked isotropic turbulence.

Fig. 1. Diagram of shocktube inlet condition. Data recorded from the isotropic turbulence simulations is an inlet condition for the shocktube domain, which is used to study shock-turbulence interactions.

[1] Lee, S. et al., *J Fluid Mech* **251,** 533 (1993).
[2] Jamme, S. et al., *Flow, Turbulence and Combustion* **68,** 227 (2002).
[3] Larsson, J. and S.K. Lele, *Phys Fluid* **21**, 126101 (2009).
[4] Livescu, D. et al., LANL Technical Report LA-CC-09-100 (2009).
[5] Petersen, M.R. and D. Livescu, *Phys Fluid* **22**, 116101 (2010).
[6] Lundgren, T., *Center for Turbulence Research Annual Research Briefs*, pp 461 (2003).
[7] Rosales, C. and C. Meneveau, *Phys Fluid* **17**, 095106 (2005).
[8] Jamme, S. et al., *AIAA* paper 2005-4886 (2005).

What is UQ?

Clint Scovel, CCS-3; Houman Owhadi, Tim Sullivan, Mike McKerns, Michael Ortiz, California Institute of Technology

We describe how Uncertainty Quantification (UQ) is currently ill-posed. We then describe the LANL and CalTech opening gambit, Optimal Uncertainty Quantification (OUQ), to resolve this situation and develop a science of UQ. When applied to a problem of seismic safety of structures, we show how the OUQ framework produces rigorous guarantees on the probability of failure of the structure as a function of the magnitude of the earthquake size. Finally, we mention that this framework can be used to develop a framework for UQ for modeling, which, along with a theory of validation, certification, and extrapolation, produces a rigorous scientific theory of certification without testing.

While everyone agrees that Uncertainty Quantification (UQ) is a fundamental component of objective science, it appears not only that there is no universally accepted notion of the objectives of UQ, there is also no universally accepted framework for the communication of UQ results. In particular, most people do not understand what UQ is, and among those that do, there is much disagreement. This dilemma has its origins, in part, in the fact that UQ, as it currently stands, has all the symptoms of an ill-posed problem–e.g. Oreskes et al. [1] assert that validation is impossible, and [2,3] describe rigorous methods for it. Moreover, often the appropriate problem is not validation but certification–using a model to assess the performance of a physical system.

Indeed, it appears that UQ is currently at the stage where probability theory was before its rigorous formalization by Kolmogorov. In an effort to resolve this situation we seek foundations for UQ. For example, in linear algebra we have vector spaces, linear transformations, the Jordan form, the spectral theorem, and the QR algorithm. Calculus has integrals and derivatives and the fundamental theorem of calculus. Linear Programming has linear programming problems and the fundamental theorems of linear programming, giving rise to the simplex algorithm. If UQ is going to be a science, then the following questions naturally come to mind:

- What are the fundamental objects of UQ?

- What are the fundamental theorems of UQ?

One important application of these ideas for the modeling community is: "What are the fundamental theorems of validation and certification?"

In an effort to resolve this situation, LANL has teamed with the Predictive Science Academic Alliance Program (PSAAP) at CalTech in an opening gambit. Our approach, which we call Optimal Uncertainty Quantification (OUQ), is very simple and follows John Dewey's [4] assertions that "a problem well formulated is half solved" and "without a problem, there is blind groping in the dark." That is, quantitatively express the problem to be solved, while quantitatively stating the assumptions being made. The main results of [5] describe OUQ optimization problems that, if solved, provide rigorous optimal solutions to UQ problems. Moreover, we describe finite dimensional reductions for a large class of these OUQ Optimization Problems. With the advent of ever more powerful computing platforms, such as the upcoming exascale systems, solutions to many of these OUQ problems may now be within our reach. Nick Hengartner at LANL tells us that this situation is similar to that experienced by the Bayesians. Namely, the Bayesian framework was not used until computers were powerful enough to compute the posteriors.

Let us apply the OUQ framework to the problem of the Seismic Safety Assessment of the structure in Fig. 1. As is standard in the seismic engineering community, we say that the structure is safe if the displacements of all the members of the truss system are less than the corresponding yield strains. Moreover, we use a standard model of the response of the truss system to ground acceleration and assume this model to be a correct representation of reality. When an earthquake hits at a fixed distance R and magnitude M on the Richter
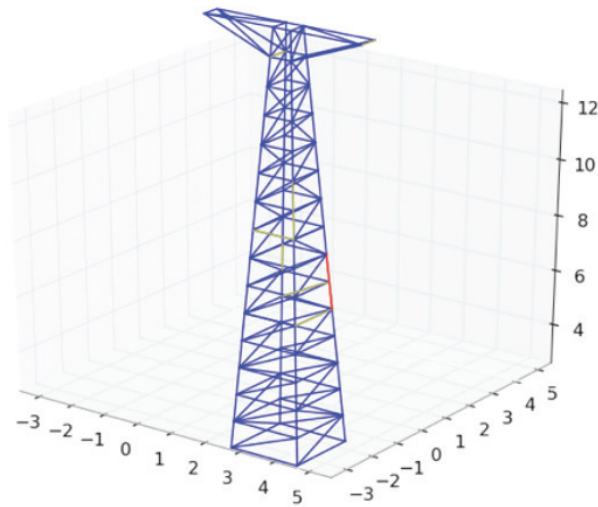
Fig. 1. Truss structure

scale, we represent the impulse structure of the earthquake in a standard way as a series of boxcar impulses with the magnitudes and timings of the boxcars being random variables. Moreover, the transfer function, defined by the structure of the earth, is also defined by some random components. The result of the earthquake is that the truss structure responds according to the response model, with the ground acceleration as input; the ground acceleration is determined by the random transfer function convolved with the random impulse model of the earthquake. So far, this is the standard methodology. Where our approach differs is that in the standard approach the random components are specified in terms of a large set of well-recognized distributions, such as Gaussians, by specifying their means and variances. We instead specify an independence structure and some simple bounds on the ranges and first moments of these random inputs. Figure 2 describes the optimal probability of failure as a function of the magnitude M for such an assumption set. The curve does not represent the probability of failure of the system, but an optimal upper bound on the probability of failure given this simple set of assumptions. Namely, there is a random model that satisfies these same assumptions but for which its probability of failure is at the value of this curve. The lower optimal bound is identically zero.

Now you may say "I don't like your assumptions!"—tell me what assumptions you do like and we will input them into our methodology and provide you with an optimal estimate of the probability of failure.

So how does this affect the modeling community? In [6] we are using this methodology to develop a comprehensive framework for uncertainty quantification for modeling, which includes, in addition to a comprehensive framework for validation, certification, and extrapolation, a scientifically rigorous approach to the problem of certification without testing.
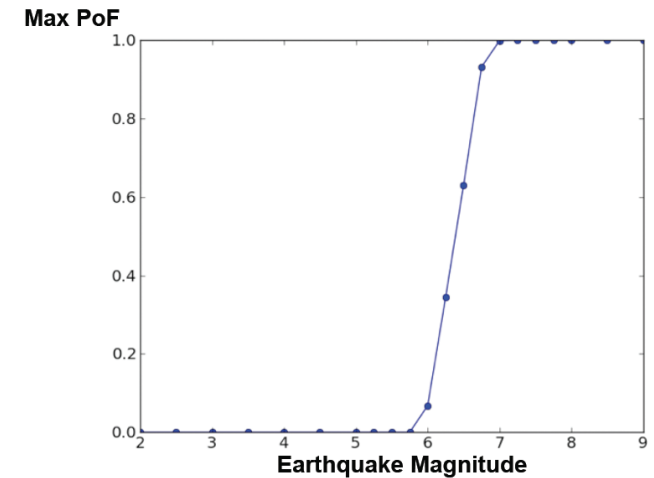
Fig. 2. Optimal upper bound on probability of failure –given the assumptions

[1] Oreskes, N. et al., *Science* **263**, 641; http://dx.doi.org/10.1126/science.263.5147.641 (1994).
[2] Lucas, L.J. et al., *Comput Meth Appl Mech Eng* **197,** 4591; http://dx.doi.org/10.1016/j.cma.2008.06.008 (2008).
[3] Scovel, C. and I. Steinwart, LANL Technical Report LA-UR-10-02355; *J Complex* **26**, 641 (2010).
[4] Dewey, J., *Logic-Theory of Inquiry,* Henry Holt and Company (1938).
[5] Owhadi, H. et al., "Optimal Uncertainty Quantification," *SIAM Review*, resubmittal under review; http://arxiv.org/pdf/1009.0679v2 (2011).
[6] Scovel, C., "Uncertainty Quantification for Modeling," unpublished results (2012).

Direct Numerical Simulation of Tilted Rayleigh-Taylor Instability

Tie Wei, Daniel Livescu, CCS-2

Rayleigh-Taylor instability (RTI) with tilted initial interface is investigated using Direct Numerical Simulation (DNS). Due to the novel nature of this flow and its importance in turbulence modeling, our simulations are also being proposed as a standard test problem for LANL codes. Our DNS results compare well with laboratory experiments for appropriately chosen initial conditions, both qualitatively and quantitatively. In general, this is a non-trivial task as there is a long-standing discrepancy between the experimental and previous numerical results of RTI-related problems. We find that the global motions in tilted RTI can be decomposed into three components: 1) rotation (overturning) of the interior interface, 2) mixing of the interior interface, and 3) development of side-wall bubble/spike. The fundamental dynamics and scaling of each component have been determined. The effects of initial conditions have also been identified.

Rayleigh-Taylor instability (RTI) is an interfacial instability that occurs when a high-density fluid is accelerated or supported against gravity by a low-density fluid. This instability is of fundamental importance in a multitude of applications, ranging from fluidized beds, oceans, and atmosphere, to inertial confinement fusion (ICF) and supernovae. Because of both its scientific and practical importance, RTI has been subjected to intense research over the last 50 years. However, previous studies, experimental or numerical, have been focused on the canonical RTI in which the perturbations are on a perfectly horizontal plane. On the other hand, in most real world applications, the initial perturbations are rarely on a perfectly horizontal plane.

In tilted RTI, the inclination of initial interface causes a large-scale overturning motion in addition to the buoyancy-driven instability. As a result, turbulence in tilted RTI is generated through two mechanisms: buoyancy and shear. The overturning of the interface also causes the mean flow to be 2D, instead of 1D as in the canonical RTI. These features make tilted RTI an excellent modeling test, and also an excellent way to study the physics of turbulence production, by contrasting the two common turbulence generation mechanisms.

Previous researches on tilted RTI have been limited to the experimental [1–3]. Here we present the first extensive studies of tilted RTI using direct numerical simulation (DNS). The DNS code we use is the CFDNS developed by Livescu et al. [4]. Numerical details and verification and validation of the code are referred to in [4,5]. Simulation parameters were chosen to match those of Rocket-Rig experiments [1]. Figure 1 shows the evolution of tilted RTI from a typical DNS simulation. Acceleration is in the vertical direction (x-direction), pointing downwards. Heavy fluid (1890 kg/m^3) is sitting on top of the light fluid (660 kg/m^3), and the Atwood number is about 0.48. The mean acceleration is about $35g_0$. Initial perturbations are on an interface tilted 5.77 degrees away from the horizontal plane (y-direction). Periodic boundary conditions are used in the z-direction (Fig. 1).

As time progresses from left to right across Fig. 1, the angle between the interior interface and horizontal plane grows continuously, until the interior interface finally becomes perpendicular to the horizontal plane. In the meantime, small perturbations on the interior interface grow due to buoyancy as in the canonical RTI. As shown in Fig. 1, bubbles and spikes of various sizes develop along the interior interface, and interact with each other. The overturning of the interior interface also causes an increase of shear within the interface, resulting in another mechanism of instability development. In contrast, the instability next to the sidewall develops quite differently from that of the interior domain. Due to the blockage of the wall, the instability next to the sidewall develops into a single "quasi-2D" bubble/spike. The growth in these regions is through a new mechanism, discovered in a separate study, through complex vortical interactions [5]. In summary, the global motions in tilted

Fig. 1. Three-dimensional view of tilted RTI evolution. From left to right: 0 ms, 26 ms, 36 ms, and 46 ms. Heavy fluid (red color) sits on top of light fluid (blue color).
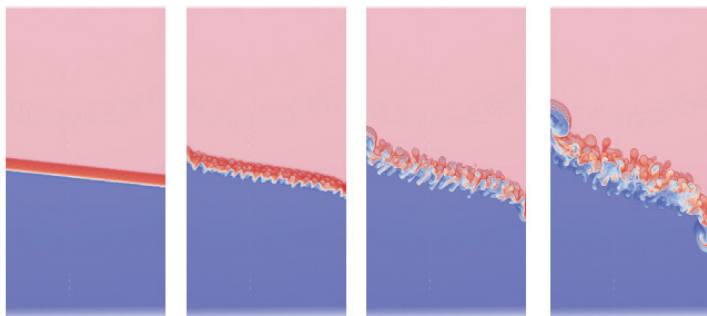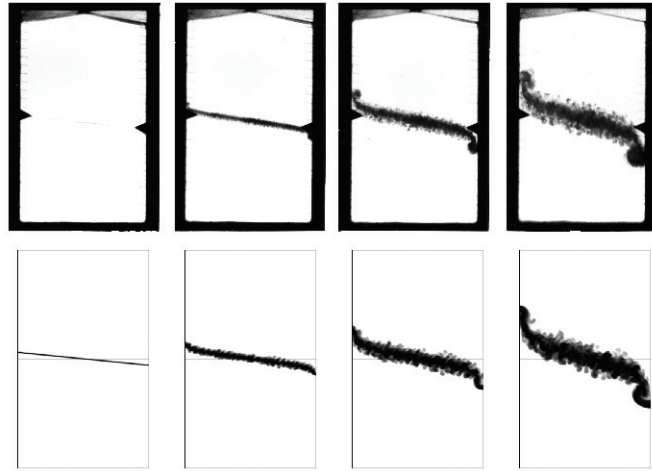
Fig. 2. Comparison between experiment (Rocket-Rig experiment number 110) and DNS. From left to right: 0 ms, 26 ms, 36 ms, and 46 ms.

RTI can be decomposed into three components: 1) rotation (overturning) of the interior interface, 2) mixing of the interior interface, and 3) development of the side-wall bubble/spike. The underlying dynamics and scaling of each component are different [6].

Our DNS results are compared, qualitatively and quantitatively, with laboratory experiments in Figs. 2 and 3. In Fig. 2, images at four instants (0 ms, 26 ms, 36 ms, and 46 ms) are compared between DNS and Rocket-Rig experiment number 110. As the figure shows, our DNS results agree well with experimental observations, including the angle and mixing of the interior interface, as well as the sidewall bubble/spike development. The agreement of the interior interface rotation can be better seen in Fig. 3, which compares the angle of the interior interface between DNS and experimental data from two independent sources [1,3]. Nevertheless, unlike experiments where only measurements of global quantities are available, all turbulence statistics necessary for model development or investigation of the turbulence physics are available from DNS. As shown in Fig. 3, the proper time scale for the rotation of the interior interface is $\sqrt{L_h / g}$, where L_h is the horizontal domain size and g is the acceleration.

In DNS, the initial conditions are precisely specified, making DNS the ideal tool to study the effects of initial conditions. The related parameters we have examined include the initial perturbation Reynolds number, the spectrum shape and amplitude of initial perturbations, and the initial tilting angle. The initial conditions have different roles in the three components of the global motions. For example, we find that the rotation of the interior interface is strongly influenced by the initial tilting angle, but not by the other parameters. In contrast, the mixing within the interior interface is strongly influenced by the initial perturbation Reynolds number.
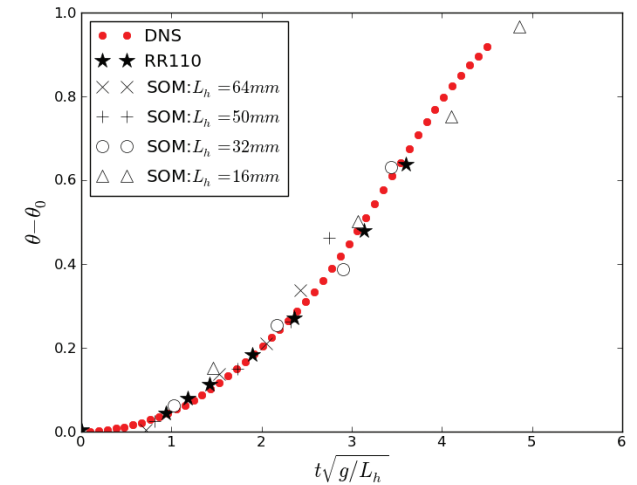
Fig. 3. Comparison of the interior interface angle between experiments and DNS. Experimental data are from two sources: Rocket-Rig experiments (RR110) [1] and SOM experiments [3]. In SOM, experimental apparatus of four horizontal domain sizes (16 mm, 32 mm, 50 mm, and 64 mm) were used.

[1] Smeeton, V.S. and D.L. Youngs, Atomic Weapon Establishment Report No. 35/87 (1987).
[2] Andrews, M. and D.B. Spalding, *Phys Fluid* **2,** 922 (1990).
[3] Ptizyna, N.V. et al., *Proc 4th International Workshop Phys Compressible Turbulent Mixing,* 129-143 (1993).
[4] Livescu, D. et al., LANL Technical Report LA-CC-09-100 (2009).
[5] Wei, T. and D. Livescu, *Phys Flui,* under review (2012).
[6] Wei, T. and D. Livescu, *J Fluid Mech,* in preparation (2012).

Los Alamos National Laboratory Associate Directorate for Theory, Simulation, and Computation (ADTSC) LA-UR 12-20429

Astrophysics and Cosmology

The work described in this section involves collaborations throughout the Theory, Simulation, and Computation Directorate and across the Laboratory. These articles represent only a small example of the great breadth and broad interest in Astrophysics and Cosmology at LANL.

The section covers subjects that span from magnetohydrodynamics turbulence to all the exotic objects in the universe–including quasars, supernovae, black holes, and neutron stars. These research efforts are both excellent examples of the high quality basic research conducted in the Directorate and indications of the overlap with programmatic activities of critical importance to fulfilling the Laboratory's nuclear security mission.

The Two States of Star Forming Clouds: AMR MHD Simulations of Self-Gravitating Turbulence

David C. Collins, T-2; Alexei G. Kritsuk, University of California San Diego, Hui Li, T-2

Star formation is one of the most fundamental unanswered problems in astrophysics. In order to build a more complete picture of star-forming clouds in the galaxy, we performed a suite of high-resolution adaptive mesh refinement (AMR) magnetohydrodynamics (MHD) simulations of self-gravitating supersonic turbulence. The simulations used large (512^3) root grids and four levels of refinement for an effective resolution of 8192^3. Mach 10 turbulence with a variety of magnetic field strengths was used. We find that the effects of self-gravity bifurcate the cloud into two distinct states: a low-density turbulent state, and a high-density self-gravitating state that can be described by an ensemble of self-similar spheres. This improved description of the statistical properties of star-forming clouds will impact star formation models, which to date have dealt with turbulence alone. Magnetic fields decrease compressibility of the gas and prevent sub-Alfvenic gas from collapsing. This can be used to explain the somewhat conflicting observations of the role of magnetic fields in the galaxy.

Fig. 1. Projected density in a subset of the domain in the mid-field strength simulation. Prestellar cores are seen in yellow in the center, collected along a filament formed form the intersection of shocks. Vectors show strength and direction of the magnetic field, showing that the prestellar objects are formed in regions where the field is weak and not aligned perpendicular to the shock.
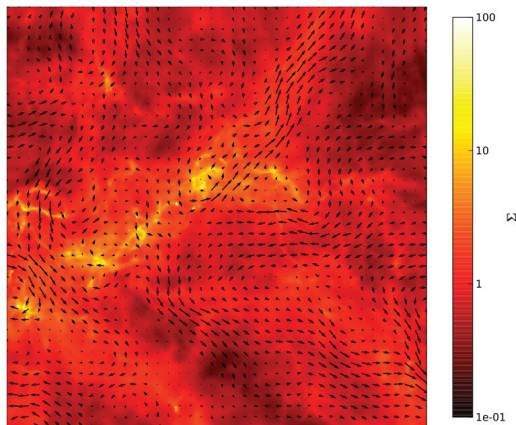
Star formation rates are observed to be significantly lower than expected from free fall alone. The two primary physical mechanisms for delaying the collapse are turbulence and magnetic fields, though the relative importance of these two effects has been hotly debated over the past fifty years [1,2]. Molecular clouds, the birthplace of nearly all stars, are observed to be both supersonically turbulent and magnetized. Unfortunately, numerical simulation of these clouds is computationally challenging because the proper resolution of turbulence requires high-resolution simulations, and the collapse involves enormous dynamic range. The advent of adaptive-resolution magnetohydrodynamics (MHD) methods [3] and the increase in computing power have allowed us to overcome these numerical restrictions and examine these systems in better detail.

We began three simulations with uniform density and magnetic fields, with a resolution of 512^3. The three field strengths were chosen so that they had initial thermal-to-magnetic pressure ratios, $\beta_0 = 0.2$, 2, and 20. The boxes were stirred at fixed resolution by a large-scale driving pattern until a statistically steady state was achieved, with a mean Mach number of 9. Then gravity and adaptive mesh refinement (AMR)

were introduced, and the simulation was evolved for 0.6 times the free-fall time at the mean density. In this time, the low-density gas was relatively unaffected by gravity, while the gas that was compressed by shocks during the driving phase was further collapsed into material that approximates self-similar collapse solutions developed earlier. Figure 1 shows a projection through one of the simulations at the end of the collapse phase. This illustrates the filamentary structure that accompanies many observed star-forming regions. The vectors show the direction and magnitude of the magnetic field. The collapsing region coincides with reduced magnetic-field strength, while the region where the field aligns with the filament is supported against collapse by magnetic pressure.

The probability density function (PDF) of supersonic isothermal turbulence has been shown to follow a lognormal distribution [4]. This lognormal has been used to predict a number of properties of star-forming clouds, such as the initial mass function of stars and the star formation rate [2,5]. While magnetized gas is not strictly lognormal, we find that the approximation is reasonable. We also find that the addition of self-gravity introduces an excess of high-density gas, which forms a powerlaw tail. The slope of this powerlaw can be understood to be the product of self-similar collapse of the gas. Figure 2 shows the density PDF, $V(\rho)$, for each simulation ($\beta_0 = 0.2$ in red, $\beta_0 = 2$ in
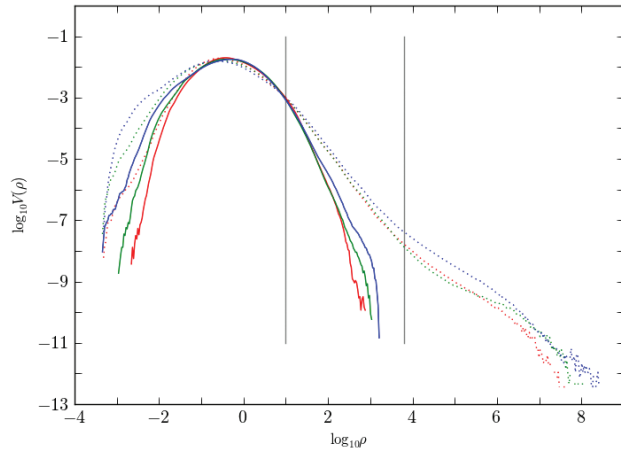
Fig. 2. Density probability distribution function for the early turbulent gas (solid lines) and the late time collapse gas (dotted lines), with red, green and blue lines showing the β_0=0.2, 2, and 20 respectively. Grey lines separate the low-density turbulent state (left) from high-density collapsing state (middle) and very-high-density gas that is numerically unresolved. Turbulence generates a lognormal density PDF, while gravity generates a powerlaw.

green, and β_0 = 20 in blue) at the beginning (solid lines) and the end (dotted lines) of the self-gravitating phase. The figure additionally shows two grey lines that divide the gas into three sections: low-density turbulent gas (left section), high-density self-gravitating gas (center section), and very-high-density gas that is numerically unresolved (right section). This gas is qualitatively interesting, but due to numerical inaccuracies is not quantitatively reliable. For the early snapshot, a lognormal is seen in all three simulations. The late snapshot shows the development of the powerlaw tail at high density, while retaining the lognormal shape at low density. Self-similar collapse solutions have the form of a powerlaw density distribution, $\rho \propto r^n$. This in turn gives a volume PDF $V(\rho) \propto \rho^m$, where $m = \dfrac{3}{n}$.

We find that the slope m increases with decreasing field strength, with m=-1.8 for the strongest mean field strength, and m=-1.64 for the weakest.

Magnetic fields are extremely difficult to observe in space, and we often must rely on circumstantial evidence to infer the strength of the magnetic field. Recent observations have shown that velocity scaling is anisotropic in some molecular clouds [6]. This is expected in strongly magnetized turbulence, as eddies are elongated in the direction of the mean magnetic field, which indicates that molecular clouds are magnetically dominated. However, this anisotropy decreases as the density increases. At stellar densities, the direction of outflows from young stellar objects is uncorrelated with the local magnetic field [7]. These observations indicate that, at some point, collapsing objects must decouple from the mean magnetic field. This behavior is also seen in, and can be explained by, our simulations. We find that the high-density state is super-Alfvenic for all values of the mean field. Figure 3 shows the ratio of dynamic pressure to magnetic pressure, β_{dyn} versus density

for the strong field case. While the low-density gas is evenly distributed around β_{dyn}=1, the high-density gas is dynamically dominated by a factor of 100. This is due to the fact that the magnetic field reduces the compressibility of the gas, thus the density in shocked material, and the dynamic pressure must overcome magnetic pressure in order to increase the density of gas to a point at which it begins to collapse.
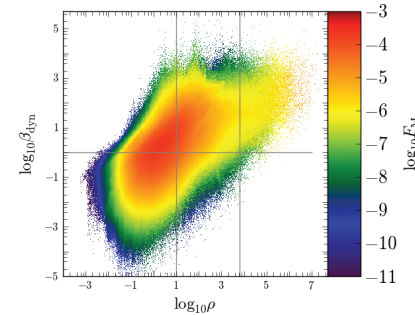
Fig. 3. Dynamic-to-magnetic pressure ratio, β_{dyn}, versus density for the strong field simulation after gravity has acted. Vertical grey lines separate turbulent (left), collapsing (middle) and unresolved (right) gas. Color shows mass fraction. Nearly all high-density gas is dominated by kinetic energy, even though the mean field is strong, similar to what is observed in nature.

[1] Mouschovias, T.C., *Astrophys J* **207,** 141 (1976).
[2] Padoan, P. and A.A. Nordlund, *Astrophys J* **576,** 870 (2002).
[3] Collins, D.C. et al., *Astrophys J Suppl* **186,** 308 (2010).
[4] Vazquez-Semadeni, E., *Astrophys J* **423,** 681 (1994).
[5] Krumholz, M.R. and C.F. McKee, *Astrophys J* **630,** 250 (2005).
[6] Heyer, M.H. and C.M. Brunt, arXiv:1110.0808 (2011).
[7] MÄnard, F. and G. DuchÉne, *A&A* **425,** 973 (2004).

Modeling Spectra and Lightcurves from Supernovae

Lucille H. Frey, XTD-6/University of New Mexico;
Wesley P. Even, XTD-6; Daniel J. Whalen, T-2/Carnegie Mellon
University; Christopher L. Fryer, CCS-2; Aimee L. Hungerford,
XTD-6; Christopher J. Fontes, XCP-5; James P. Colgan, T-1

Massive stars at the end of their lives release huge amounts of energy in supernova explosions that can be detected across cosmological distances. Even if prior observations exist, such distances make supernova progenitors difficult to identify. New real-time surveys, like the Large Synoptic Survey Telescope (LSST) and the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS), may detect supernovae (SNe) at very early times, giving us a rare view of these short-lived stars immediately after core collapse.

When the radiation-dominated shock wave from a core collapse approaches the stellar surface, the optical depth of the plasma ahead of the shock decreases until the radiation can escape in a burst. Gaining energy from the shock's kinetic energy via Compton scattering, this burst can be observable as an ultraviolet (UV) or X-ray flash, lasting for seconds to hours. If a dense stellar wind is present, shock breakout (SBO) can occur at the edge of the wind. Occurring days or weeks before the optical light from radioactive decay peaks, shock-breakout radiation can be used to determine the radius of the progenitor star or its recent mass loss history. This early detection of supernovae (SNe) allows spectra and luminosity observations to be made during the rising phase of the light curve as well as the decline.

Shock breakout was predicted nearly 40 years ago [1,2], but has only recently been observed as an early peak in the UV light curve [3,4], as infrared (IR) echoes in the Cassiopeia A SNe remnant [5], and directly as an X-ray burst in GRB 060218/SN2006aj [6] and XRO 080109/

Fig. 1. Hydrogen opacity at 0.5 eV for a range of densities.

SN2008D [7]. Simulations can help match our theoretical understanding of core-collapse SNe and SBO to these observations. Analyses of these observations (for example, [4, 7-9]) have considered various progenitor stars and stellar environments, considering whether SBO occurs near the stellar surface or in a wind or shell. In all of these environments, SBO is treated as a single wavelength-independent event for which a single radius, temperature, and time scale can be observed or calculated. This assumes a gray opacity, which comparisons to monochromatic opacities can show to

be invalid (see Fig. 1). We have developed a new code, Spectrum, to calculate spectra and lightcurves with monochromatic opacities and allow us to study SBO as a wavelength-dependent phenomenon. We describe the codes used for the SNe simulations presented here, the Spectrum code, and some preliminary results from this pipeline.

The results presented here are from a 1D simulation of a 23-solar-mass progenitor [10] that was evolved with the TYCHO stellar evolution code as a binary star. The H envelope is lost at the base of the red giant branch and the star evolves as a Wolf-Rayet star, ending as a 6.4-solar-mass star that explodes as a Type Ib SNe. A 1D code simulates core-collapse and the launch of the shock wave, and the central neutron star is replaced with a hard reflective boundary and a gravitational potential. After nuclear burning is finished, the output is mapped into RAGE (Radiation Adaptive Grid Eulerian), a multidimensional, multispecies Eulerian adaptive mesh refinement (AMR) radiation hydrodynamcs code [11]. RAGE uses a cell AMR scheme, allowing adjacent cell sizes to vary by no more than one level. We use two-temperature physics in which matter and radiation are coupled but can be at different temperatures. Each of up to 30 species is evolved with its own continuity equation.

Spectrum is a post-process code that takes density, temperature, velocity, and material mass fraction profiles as input and creates spectra for a wide wavelength range [12]. It uses monochromatic opacities from the LANL OPLIB database [13], which contains single-element opacities for local thermodynamic equilibrium (LTE) conditions. These opacities were accessed through the TOPS code for temperature and density pairs for 14,900 energies. Spectrum maps 1D data onto a 2D grid with a
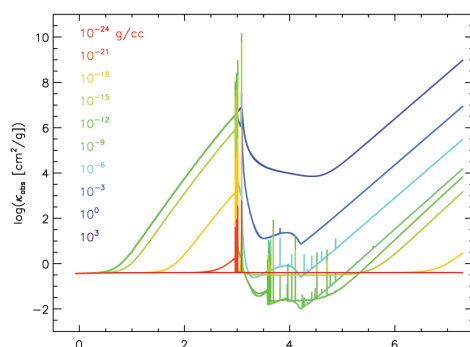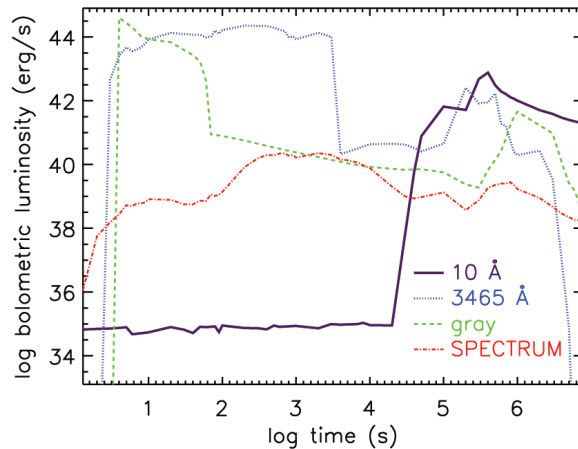
Fig. 2. Bolometric lightcurves using a blackbody approximation at a tau=1 radius calculated with monochromatic opacities at two wavelengths and a gray helium opacity, and the bolometric lightcurve from Spectrum.

specified number of zones in angle, uniformly spaced in cosine. Each of these fluid elements is assumed to be in LTE with its surroundings and to emit as a blackbody. This blackbody luminosity for each wavelength is attenuated along a line of sight through the material between that fluid element and the observer by integrating over the absorption opacity. These luminosity calculations also include the effects of Doppler shifting, time dilation, and limb darkening. Lightcurves are created by integrating over each spectra over a given wavelength band.

A variant of Spectrum continues to be developed to analyze dumps from the Advanced Simulation and Computing program (ASC) RAGE and Cassio codes that are simulating experiments on the National Ignition Facility, the Omega laser, and the Z-machine at SNL. This Spectrum code variant is able to calculated the exact diagnostics in these experiments, allowing a more direct comparison between simulation and experiment and allowing LANL scientists to do more rigorous verification and validation within its experimental program.

The radius (and time) at which shock breakout occurs is dependent on the opacity of the outer layers of the star and any surrounding wind or ejecta. Monochromatic opacities can vary by many orders of magnitude with wavelength, so a single radius and temperature cannot accurately describe this phenomenon. To study this phenomenon we used Spectrum to calculate the tau=1 surface (where the material becomes optically thick) using monochromatic opacities for a range of wavelengths, as well as a gray helium opacity. Using the Stefan-Boltzmann law, we calculated a bolometric luminosity using the radius and temperature at this surface for several wavelengths and the gray case, and compared these to the bolometric luminosity from the full Spectrum calculation (see Fig. 2). The single temperature and radius blackbody approximation overestimates the total luminosity, since no attenuation is present, and does not match the rise time or rate observed for a monochromatic

Fig. 3. Lightcurves for Swift UV/ Optical bands created with Spectrum.

calculation. Creating lightcurves for multiple wavelength bands using Spectrum shows how the observed peak luminosities, rise time, and duration changes with wavelength (see Fig. 3). A future paper will describe SBO in more detail for single and binary core-collapse SNe progenitors and explore the implications of the wavelength dependence of SBO for comparing simulations to observations.

[1] Colgate, S.A., *Astrophys J* **187,** 333 (1974).
[2] Klein, R.I. and R.A. Chevalier, *Astrophys J Lett* **223,** 109 (1978).
[3] Schawinski, K. et al., *Science* **321,** 223 (2008).
[4] Gezari, S. et al., *Astrophys J Lett* **683,** 131 (2008).
[5] Dwek, E. and R.G. Arendt, *Astrophys J* **685,** 976 (2008).
[6] Campana, S. et al., *Nature* **442,** 7106 (2006).
[7] Soderberg, A.M. et al., *Nature* **453,** 469 (2008).
[8] Chevalier, R.A. and C. Fransson, *Astrophys J Lett* **683,** 135 (2008).
[9] Waxman, E. et al., *Astrophys J* **667,** 351 (2007).
[10] Young, P.A. et al., *Astrophys J,* **891** (2006).
[11] Gittings, M. et al., *Comput Sci Discov* **1,** 015005 (2008).
[12] Frey, C.L. et al., in preparation (2012).
[13] Magee, N.H. et al., *Astron Soc Pac Conf* **1294,** 268 (1995).

Gamma-ray Bursts from Stellar Collapse

Christopher L. Fryer, CCS-2; Daniel E. Holz, T-2;
Chris Belczynski, University of Warsaw;
Gabriel M. Rockefeller, CCS-2; Gang Shen, T-2

One of the leading models behind gamma-ray bursts postulates that these cosmic explosions are produced when a star collapses to form a black hole (BH). The subsequent accretion onto the BH produces the explosion. To understand these BH accretion disk gamma-ray burst (GRB) models, we must understand BH formation. We conducted a detailed study of BH formation, comparing our results with observed X-ray binary systems. Finally, these studies were applied to the recently observed Christmas burst.

The leading model behind long-duration gamma-ray bursts (GRB) invokes the collapse of a massive star down to a black hole (BH) [1], but the exact details of this collapse and the exact progenitor remains unknown, with a host of progenitor possibilities [2]. To differentiate these progenitors, we must understand the details of BH formation and, to do this, we must understand stellar collapse. This year, we led a series of efforts to better understand BH formation and the progenitors of GRBs.

Stars are powered by fusion in their cores. Massive stars proceed through a series of core-burning phases, where the ashes of the previous phase become the fuel for the next phase. In this manner, the core burns hydrogen to helium, helium to carbon/oxygen, carbon to silicon, and silicon to iron. At this point, fusion of iron does not yield any additional energy. As the iron core increases in size, it ultimately becomes so massive that thermal and degeneracy pressure can no longer support the core and it collapses. The potential energy released in the collapse of the iron core of a massive star down to a neutron star (NS) is what powers type Ib/c and type II supernovae. However, in some stars the released energy is unable to drive a strong supernova explosion. In these cases, the proto-NS collapses to form a BH.

The fate of the collapsing star ultimately depends on the structure of the star. This year, we developed an analytic framework with which we are able to analyze stellar models and determine the fate of the collapsing stars and the final remnant mass of the collapsed core. By studying a large set of stars at collapse, LANL scientists developed correlations between the initial mass of a star and its final compact remnant mass after collapse [3]. With these correlations for both single and binary star systems, we were able to develop distributions of NS and BH masses in the galaxy that can be directly compared to observed distributions.

BH mass measurements rely on a complex combination of challenging observations of X-ray binaries (in quiescence, if they are transient) and of modeling of photometric and spectroscopic data. The uncertainties associated with these measurements are more significant than in the case of NSs. Early analysis argued that the measurements are consistent with a relatively narrow mass distribution [4] around seven solar masses. Recent analyses [5,6] have used the expanded current samples of BH measurements in both Roche-lobe-overflow and wind-driven X-ray binaries and proposed distributions to fit the observations (without quantitative consideration of selection effects). The statistically favored models have mass distributions that extend to high masses (~15–20 solar masses), depending on whether the wind-fed systems with more massive BHs (which are more uncertain) are included. Both studies conclude that there is clear evidence for a low-mass gap in the distribution, with no remnants found in between the maximum NS mass of about two solar masses and four to five solar masses. However, this field is still in a state of flux as new observations appear: observations of IGR J17081-3624 suggest that at least some low-mass BHs exist [7].

Fig. 1. Number of compact remnants produced as a function of remnant mass. Those systems above two to three solar masses are BHs. The top panel shows the observed systems, the bottom panel shows our simulated results.
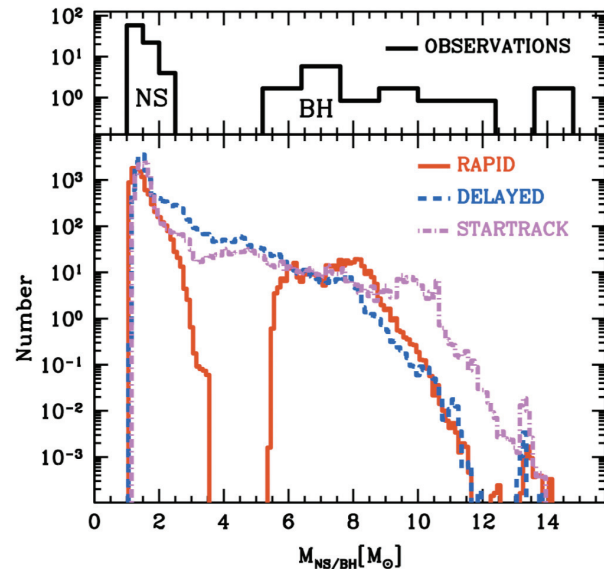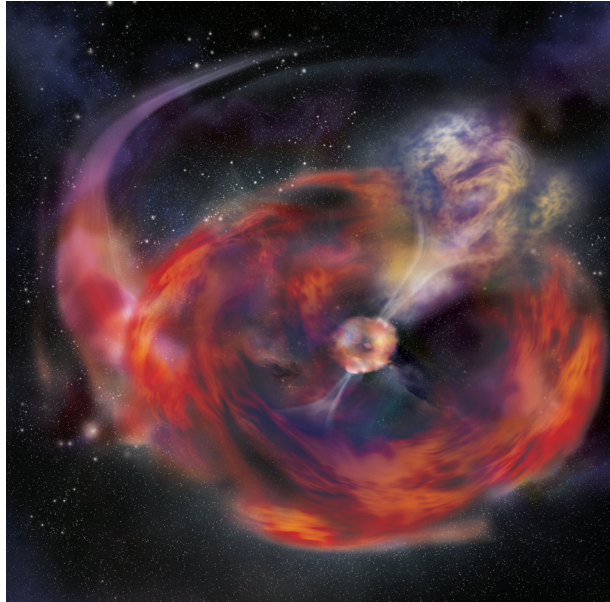
Fig. 2. Artist's drawing of the Christmas burst showing the surrounding debris produced in our merger model.

These observed mass distributions are a select sample of the BHs in the universe. This sample is limited to those binary systems in which, after BH formation, the system is close enough to transfer mass from the companion to the BH, producing X-ray emission. These so-called "X-ray binaries" make up our observed mass distribution. To directly compare to our theoretical correlations with observations, we ran population synthesis models that evolved the stellar binaries until they became X-ray binaries. A comparison of theory and observation is shown in Fig. 1. Although the theory distribution still allows some systems within the range of two to four solar masses, we note that the recent observation of IGR J17081-3624 shows that some systems in the gap exist.

Although the primary factor in determining the fate of the star is the structure of the pre-collapse star, we have many uncertainties in the collapse that we must understand before we fully understand its fate. One of the most important is the role of the equation-of-state. With our computing time on the Turquoise network and using the SNSPH code [8], we were able to model the collapse of a 60-solar-mass start to times well beyond anything done to date. We are currently coupling the new Shen equation-of-state to the SNSPH code and we conduct detailed comparisons of this long run with a new run using this different equation-of-state. These comparisons will allow us to better quantify the importance of the equation-of-state on BH formation.

All of this work allows us to much better understand the BH-forming phenomenon that we observe in nature. We conclude with a discussion of one such example. On December 25, 2010, the Swift Burst Alert Telescope detected a GRB, one of the longest GRBs ever observed by Swift with a duration of over 2000 s (using the time in which 90% of the gamma-ray energy is released). The most surprising feature of this "Christmas burst" is the spectral energy distribution (SED) of its afterglow [9]. The X-ray SED is best modeled with a combination of an absorbed power-law and a blackbody (BB). The UV/optical/Near Infra-Red (UVOIR) SED can be modeled with a cooling and expanding BB model until 10 days post-burst, after which we observe an additional spectral component accompanied by a flattening of the light curve. This behavior differs from a normal GRB where the SED follows a power-law due to synchrotron emission created in shocks when the jet hits the interstellar medium.

Our studies of BH formation allowed us to quickly develop a plausible progenitor for this burst and we were able to develop a model that explained the observed features of this explosion. Our model, based on earlier work by Fryer and colleagues in the 1990s [10,11] invoked a compact object/main sequence merger just prior to the outburst. In essence, this model is an extreme version of the X-ray binaries we studied to understand the compact mass distribution. Instead of slowly accreting matter from its compact object, the compact object spirals into its companion's core. The subsequent accretion drives the powerful GRB explosion. The ejecta from the merger produce a shell responsible for many of the peculiar features of this explosion.

[1] Woosley, S.E., *Astrophys J* **405**, 273 (1993).
[2] Fryer, C.L. et al., *Astrophys J* **526**, 152 (1999).
[3] Fryer, C.L. et al., "Compact Remnant Mass Function: Dependence on the Explosion Mechanism and Metallicity," astro-ph\1110.1726 (2011).
[4] Bailyn, C.D. et al., *Astrophys J* **499**, 367 (1998).
[5] Özel, F. et al., *Astrophys J* **725**, 1918 (2010).
[6] Farr, W.M. et al., *Astrophys J* **741**, 103 (2011).
[7] Altimirano, D. et al., *Astrophys J* **742**, L17-21 (2011).
[8] Fryer, C.L. et al., *Astrophys J* **643**, 292 (2006).
[9] Thone, C. et al., *Nature* **480**, 72 (2011).
[10] Fryer, C.L. and S.E. Woosley, *Astrophys J* **502**, L9-12 (1998).
[11] Zhang, W., and C.L. Fryer, *Astrophys J* **550**, 357 (2001).

The Growth of the Stellar Seeds of Supermassive Black Holes

Jarrett L. Johnson, T-2; Daniel J. Whalen, T-2, Carnegie Mellon University; Christopher L. Fryer, CCS-2; Hui Li, T-2

The collapse of baryons into extremely massive stars with masses $>10^4$ M$\odot$ in a small fraction of protogalaxies in the early universe is a promising scenario for the origin of supermassive black holes (BH). We determine the maximum masses such stars can attain by the accretion of primordial gas before they collapse to BHs. We find that the strong ionizing radiation emitted by these stars limits their masses to ~10^3 M$\odot$ ($R_{acc}/10^{-3}$ M$\odot$ yr^{-1})$^{8/7}$, where R_{acc} is the rate at which a star gains mass. However, at extremely high accretion rates, usually found in numerical simulations of protogalactic collapse ($R_{acc} >0.1$ M$\odot$ yr^{-1}), the lifetime of the star instead limits its final mass to ~10^6 M$\odot$. We predict that unique observational signatures of rapidly accreting supermassive stars may be detected by future missions such as the James Webb Space Telescope.

Fig. 1. The maximum mass to which a star can grow at a constant accretion rate R_{acc}. Above the maximum mass $M_* \propto R_{acc}^{8/7}$ (blue curve), radiative feedback shuts off accretion because ionizing radiation from the star transfers momentum to the gas and prevents it from reaching the stellar surface. However, for extremely high accretion rates $R_{acc} > 0.1$ M$\odot$ yr^{-1}, the maximum mass is set instead by the lifetime of the star (red curve), which is shortened due to the rapid rate at which nuclear fuel must be burned in order to support the star against its ever-increasing weight. As the red curve shows, this leads to a maximum stellar mass, at the highest accretion rates, of ~1.5×10^6 M$\odot$. The dotted horizontal line delineates the mass.

The origin of ~10^9 M$\odot$ black holes (BH) in massive galaxies at redshift z~7, less than a billion years after the Big Bang [1,2], is one of the great mysteries of cosmological structure formation. In the standard Lambda cold dark matter ($\triangle$CDM) cosmological paradigm, early structure formation is hierarchical, with small dark-matter halos at early epochs evolving into ever more massive ones by accretion and mergers through cosmic time. Hence, it is generally held that supermassive black holes (SMBH) grow from much smaller seed BHs at high redshifts. The origin of these seeds, and how they reach such large masses by such early times, remains to be understood.

A promising scenario for the formation of SMBH seeds is the rapid collapse of primordial gas in protogalaxies into supermassive stars with masses >10^4 M$\odot$, which upon their deaths collapse to similarly massive BHs [3]. Such large initial masses may be necessary in order for the BHs to grow quickly enough to explain the presence of SMBHs in the very early universe [4]. Numerical simulations of this process [5–8] find that supermassive stars grow by accretion of gas at rates often exceeding ~0.1 M$\odot$yr^{-1}, orders of magnitude larger than is found for stars forming in the present-day universe.

We address a fundamental question: "What ultimately limits the growth of supermassive stars and thus sets the mass scale of seed BHs?" There are two processes that limit the mass to which such stars can grow, even at the highest accretion rates believed to take place in the early universe. The first of these is so-called radiative feedback from the star. Primordial stars, in particular, emit copious high-energy radiation that ionizes the gas surrounding them. When the high-energy photons from the star are absorbed, momentum is transferred to the gas and it receives a push away from the star. This is the process that we have found to be most effective at shutting off accretion [9]. In particular, by modeling how the dynamics of the accretion flow are altered by the radiation emitted from the growing star, we have found that the maximum mass M_* that a primordial star can obtain by accreting gas at a constant rate R_{acc} is well approximated by $M_* \sim 10^3$ M$\odot$ ($R_{acc}/10^{-3}$ M$\odot$ yr^{-1})$^{8/7}$. At higher accretion rates, the accretion flow is not stopped by the radiation from the star, and the star continues to grow; accretion at lower rates is not possible, as the radiation from the star prevents the gas from reaching the stellar surface. This maximum stellar mass is shown by the blue curve in Fig. 1. As the figure shows, the process of radiative feedback determines the final mass of stars that grow by accretion at rates $R_{acc} <0.1$ M$\odot$ yr^{-1}.

The second process that limits the growth of supermassive stars is their limited lifetime, which can be shorter than a million years, much shorter than the ~10-billion-year lifetime of most stars formed in the present-day universe. When the star exhausts its nuclear fuel its core collapses to a BH, which quickly swallows the entire star [10]. For a star to become supermassive it must therefore grow very quickly before
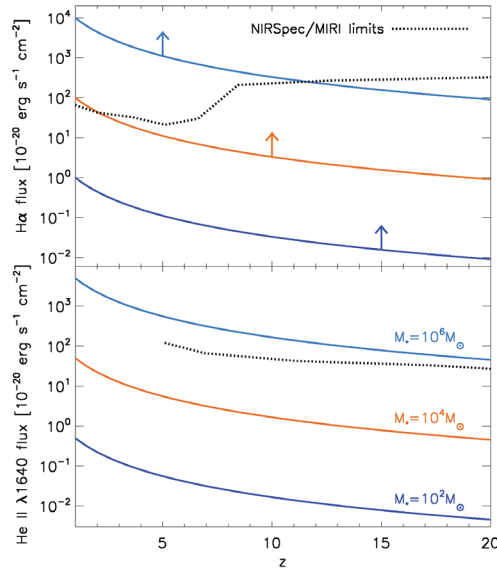
Fig. 2. The flux in Hα (top panel) and He II λ1640 (bottom panel) from rapidly accreting primordial stars, as a function of redshift z, for stellar masses of 10^2 (dark blue), 10^4 (red), and 10^6 M⊙ (light blue). The dotted black lines show flux limits for detection of these lines with the Near-Infrared Spectrograph (NIRSpec) and the Mid-Infrared Instrument (MIRI) on the JWST. NIRSpec operates at wavelengths of ≅1–5 μm, while MIRI covers the range from ≅5–28 μm. The Hα fluxes shown here are lower limits, as indicated by the arrows. While detection of Hα from stars with masses of >10^4 M⊙ may be possible out to redshifts z >10, only accreting primordial stars with masses of at least ~10^5 M⊙ are likely to be detectable in He II λ1640.

its time runs out. However, the faster the star grows, the more quickly it must burn its nuclear fuel in order to support its ever-increasing weight–the result being that the faster a star grows the shorter is its lifetime [11]. We find that this process results in a maximum possible stellar mass of ~1.5 x 10^6 M⊙ which is asymptotically approached at the highest accretion rates [9], as shown by the red curve in Fig. 1. As shown in the figure, this process sets the maximum mass of stars that grow by accretion at rates R_{acc} >0.1 M⊙yr^{-1}. As accretion rates in this range are found in cosmological simulations of the collapse of primordial protogalaxies, our findings suggest that there is nothing preventing the stellar seeds of supermassive black holes from forming with masses approaching ~10^6 M⊙.

Given that supermassive stars may indeed have been the seeds of SMBHs in the early universe, it is an important question whether or not these objects could be detected and identified observationally. In Fig. 2 we show the flux emitted from rapidly accreting primordial stars as a function of redshift z for three representative stellar masses: 10^2, 10^4, and 10^6 M⊙. In this figure we also show the flux limits for deep infrared surveys to be carried out by the James Webb Space Telescope (JWST) [12]. While the most luminous emission line predicted for primordial stars is the Lyman α line of atomic hydrogen [13], we find that this line is scattered so readily in the accretion flow that it cannot escape to be observed. Instead, this emission is likely reprocessed largely into Hα line emission from atomic hydrogen, which does escape and could be detected; shown in the top panel of Fig. 2 are lower limits for the flux in this line. In the bottom panel, we show the flux in the 1640 Angstrom emission line of ionized helium, which is predicted to be very strong from primordial stars. Together, the detection of Hα and He II 1640, along with the

non-detection of Lyman α, are unique observational signatures of rapidly accreting supermassive primordial stars. As the figure shows, these signatures could be detected by the JWST from stars with masses of at least ~10^5 M⊙. This opens up the exciting possibility that in the coming years missions such as the JWST may reveal the nature of the earliest seeds of SMBHs.

[1] Fan, X. et al., *Astron J* **125**, 1649 (2003).
[2] Mortlock, D.J. et al., *Nature* **474**, 616 (2011).
[3] Bromm, V. and A. Loeb, *Astrophys J* **596**, 34 (2003).
[4] Volonteri, M., *Astron Astrophys Rev* **18**, 279 (2010).
[5] Wise, J.H. et al., *Astrophys J* **682**, 745 (2008).
[6] Regan, J.A. and M.G. Haehnelt, *Mon Not Roy Astron Soc* **396**, 343 (2009).
[7] Shang, C. et al., *Mon Not Roy Astron Soc* **402,** 1249 (2010).
[8] Johnson, J.L. et al., *Mon Not Roy Astron Soc* **410**, 919 (2011).
[9] Johnson, J.L. et al., ArXiv e-prints, 1112.2726 (2011).
[10] Fryer, C.L. and A. Heger, *Astron Notes* **332**, 408 (2011).
[11] Begelman, M., *Mon Not Roy Astron Soc* **402**, 673 (2010).
[12] Gardner, J.P. et al., *Space Sci Rev* **123**, 485 (2006).
[13] Raiter, A. et al. *Astron Astrophys* **523**, 64 (2010).

Modeling the Ocean in an Accreting Neutron Star

Zachary J. Medin, XCP-5; Andrew Cumming, McGill University

Neutron stars provide some of the most extreme physical conditions in the universe, from the super-strong magnetic fields surrounding and threading these stars to the super-nuclear densities in their interiors. Many of these conditions cannot be recreated in the laboratory, making neutron stars an important and unique probe of fundamental physics. X-ray telescopes, such as Chandra and XMM-Newton, have allowed astronomers to make detailed observations of the surface emission from a large number of neutron stars. These observations can reveal a great deal about neutron stars, if the outer layers are accurately modeled. We present models of the ocean in accreting neutron stars–used in tandem with models of the neutron star crust and core and with X-ray observations, our models can potentially place strong constraints on the properties of these layers.

The ocean of an accreting neutron star is composed of carbon and a variety of heavier elements, and is formed by the nuclear burning of the accreted hydrogen and helium at low densities. As it is compressed to higher densities by ongoing accretion, the matter in the ocean eventually solidifies to form the crust. As the matter is compressed further, heat is released through electron captures and pycnonuclear reactions–this "deep crustal heat" flows upward into the ocean and downward into the core.

The ocean of an accreting neutron star can be observed in (at least) two ways: during an X-ray superburst, and about a week after accretion turns off in a transiently accreting neutron star. In a superburst, the carbon near the base of the ocean reaches a critical density and temperature for unstable ignition, whereupon the entire ocean burns in a runaway thermonuclear event. A large portion of the heat flows upward to the surface of the neutron star and is seen as a large X-ray flash lasting nearly a day. In a transiently accreting neutron star, as soon as accretion turns off, the neutron star begins cooling from the outside in, down to the depth of the inner crust. The top of the ocean begins cooling within a day, and the base of the ocean begins cooling after a week—therefore, observations of the change in surface emission with time in the first week of quiescence will probe the neutron star ocean temperature [1].

At the ocean-crust boundary, as the matter transitions from liquid to solid, it also undergoes chemical separation. Numerical simulations of phase transitions in neutron stars [2] have shown that as the ocean mixture solidifies, light elements (charge numbers $Z < 20$) are preferentially left behind in the liquid whereas heavy elements are preferentially incorporated into the solid. Standard models of the outer layers of an accreting neutron star, which do not include chemical separation in the ocean, cannot produce the high carbon densities and temperatures in the ocean that are required for superbursts [3]. Our models help solve both of these problems. We find that the retention of light elements in the liquid acts as a source of buoyancy that drives a continual mixing of the ocean, enriching it substantially in light elements and leading to a relatively uniform composition with depth [4]. Heat is also transported inward to the ocean-crust boundary by this convective mixing (see Figs. 1 and 2).

Cooling of neutron star transients provides an independent way to constrain chemical separation and mixing in the ocean. As the neutron star cools after an accretion episode, the ocean-crust boundary moves upward. We find that this leads to chemical separation, and then convective mixing and inward heat transport, in a manner similar to that during accretion. The inward heat transport cools the outer layers of the ocean rapidly, but keeps the inner layers hot; the result is a sharp drop in surface emission at around a week, followed by a gradual recovery as cooling becomes dominated by the crust (see Fig. 3). Such a dip should be observable in the light curves of these neutron star transients, if enough data is taken at 10 to 100 days after the end of accretion.
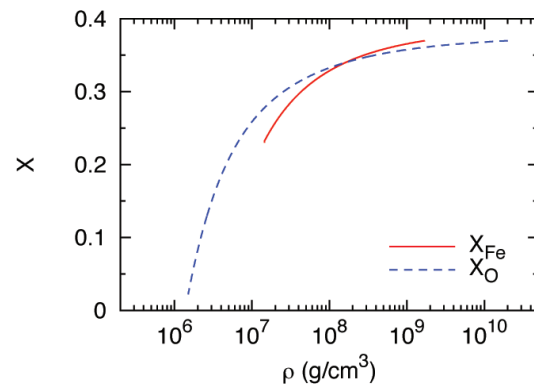
Fig. 1. Composition profile in the convection zone of an ocean composed of iron and selenium (solid red line) and oxygen and selenium (dashed blue line).
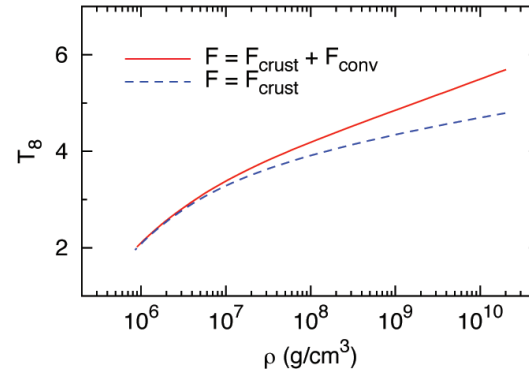
Fig. 2. Thermal profile for an ocean composed of oxygen and selenium, both when the convective flux is included in the total heat flux F (solid red line), and when it is ignored (dashed blue line).
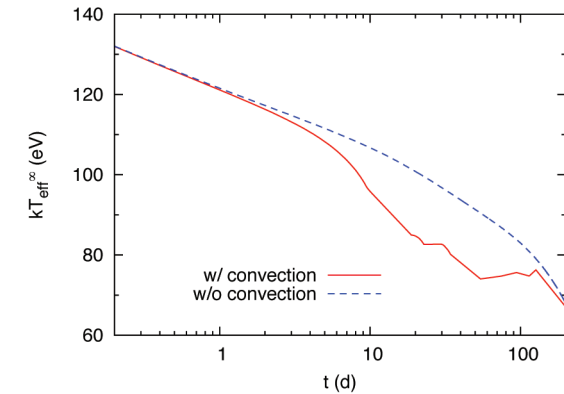
Fig. 3. Example lightcurve during the cooling of a transiently accreting neutron star, both when cooling-induced convection is included (solid red line), and when it is ignored (dashed blue line).

[1] Brown, E.F. and A. Cumming, *Astrophys J* **698,** 1020 (2009).
[2] Horowitz, C.J. et al., *Phys Rev E* **75,** 066101 (2007).
[3] Cumming, A. et al., *Astrophys J* **646,** 429 (2006).
[4] Medin, Z. and A. Cumming, *Astrophys J* **730,** 97 (2011).

Mixing in Converging Flows

Candace C. Joggerst, T-2; Anthony Nelson, University of North Carolina; Paul Woodward, University of Minnesota; Catherine Lovekin, T-2; Thomas Masser, CCS-2; Guy Dimonte, XCP-5; Christopher L. Fryer CCS-2; Praveen Ramaprabhu, University of North Carolina; Marianne M. Francois, Gabriel M. Rockefeller, CCS-2

The growth of turbulence in converging flows can affect scenarios as small as the implosion of inertial confinement fusion pellets and as large as the collapse and explosion of massive stars. Such scenarios are often challenging to model, both from a purely hydrodynamic perspective and because they often require accurate simulation of many different physical processes. We present initial results of an investigation into turbulence and mixing in converging flows, using four different codes to simulate cylindrical and spherical implosions on 2D meshes. The results of these simulations lay the groundwork for future 3D studies.

In a problem involving the implosion of a sphere or cylinder, Rayleigh-Taylor and Richtmyer-Meshkov instabilities may both be present. The Rayleigh-Taylor instability develops when the density and pressure gradients in a fluid are opposed, as when a denser fluid rests atop a lighter fluid in a gravitational field, or when a lighter fluid accelerates a denser fluid. The Richtmyer-Meshkov instability develops when a shock passes through an interface between a heavy and a light fluid.

Most simulations and physical experiments that study the details of the growth of these instabilities have been performed in planar geometry (for example, see [1]). However, Rayleigh-Taylor and Richtmyer-Meshkov instabilities also play important roles in non-planar problems. Instability growth on such disparate scales as inertial confinement fusion capsules and the interiors of pre-supernova stars may dramatically influence the evolution of these implosions [2].

Ideally, simulations of physical phenomena should minimize the presence of artifacts of the numerical method. Simulations of instabilities in radially converging flows present extra challenges beyond those present in simulations of instabilities in planar interfaces; in particular, the ideal spatial coordinate system and mesh are no longer obvious. While spherical coordinates may seem appealing for following the flow of fluid in spherical implosions, the simplicity of mesh-aligned flow is lost once turbulence develops. Spherical coordinates also introduce coordinate singularities that impose a preferred orientation on the simulation, and features may develop differently along the axis than at other points around the sphere. In multiphysics problems, such features can expand and contaminate

Fig. 1. Simulated 2D cylindrical implosions of dense shells of fluid surrounding relatively light cores, near the time of maximum compression.

regions far from the axis. Cartesian coordinates avoid any coordinate singularities and are suitable for simulating a wide variety of physical problems, but using them to simulate spherical implosions raises questions about the spatial resolution required not just for simulating curved interfaces on rectilinear meshes, but also for following the growth of perturbations imposed on those interfaces.

Youngs and Williams [3] used a Lagrange-remap code on a 3D spherical polar mesh to simulate turbulent mixing in a sector of a spherical implosion with random perturbations that were initially applied to the interface between light and dense fluids. The authors found that the width of the mixing zone shrank slightly as they increased the resolution of the mesh but that, at the finest mesh resolution they used, the mixing zone width seemed to have converged.

Do simulations in other coordinate systems, on other meshes, achieve similar levels of convergence? Specifically, what resolution would be required for a simulation of a spherical implosion on a Cartesian mesh to yield a converged measurement of the mixing zone width? Do other diagnostics of mixing and turbulence demonstrate the same degree of convergence as mesh resolution increases?

For this project, we compare results of simulations of radially converging flows from four codes: RAGE [4], CASTRO [5], FLASH [6], and multi-fluid PPM [7]. All four codes model compressible hydrodynamics on Eulerian meshes, but they differ in the details of their discretization schemes, their shock-capturing methods, limiters and steepeners, their treatment of materials in mixed cells, and their support for different coordinate systems. We assess the impact of the choice of coordinate system on the ability of these codes to follow the growth of turbulence from perturbed interfaces, and we have begun to quantify
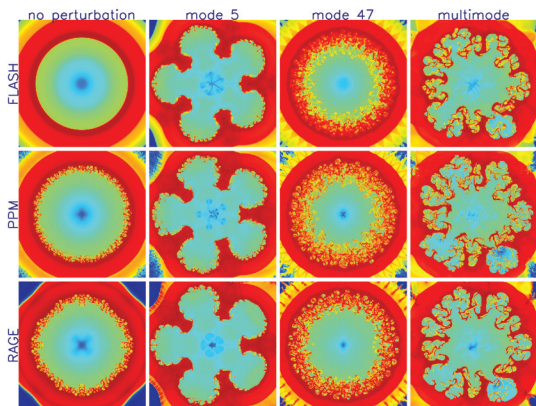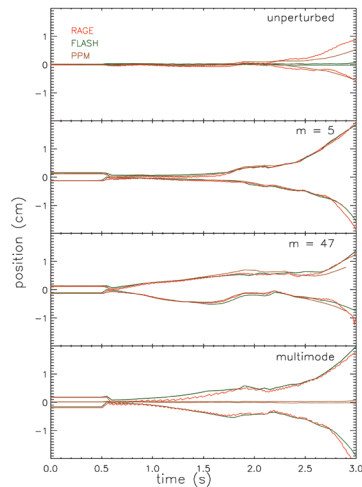
Fig. 2. Azimuthally averaged heights of bubbles of light fluid, and depths of spikes of dense fluid, as a function of time in the cylindrical implosions.

the spatial resolution required to follow perturbations with different wavelengths and initial amplitudes.

Figure 1 shows snapshots of simulated 2D cylindrical implosions of dense shells of fluid surrounding relatively light cores, near the time of maximum compression. Color represents the mass fraction of the dense fluid–the reddest regions contain unmixed dense fluid, while the bluest regions contain unmixed light fluid, and yellows and greens indicate regions where substantial mixing has occurred. The interface between light and dense fluids was either left unperturbed (left-most column) or initialized with one of three different perturbations—a long-wavelength perturbation (mode 5), a short-wavelength perturbation (mode 47), or a multi-mode spectrum.

Results from the three perturbed-interface simulations compare well across all codes. This degree of agreement is particularly interesting considering the different mesh geometries and refinement strategies employed by the codes: the RAGE calculations used a Cartesian mesh with adaptive mesh refinement, the FLASH calculations used a polar mesh with adaptive mesh refinement, and the PPM calculations used a uniform Cartesian mesh that moved inward with the flow. Differences appear at small scales–for example, in the specific pattern of secondary instabilities at the edges of the inward spikes of dense fluid in the mode-5 calculations. Using a polar mesh, FLASH is able to preserve an unperturbed cylindrical interface, while PPM and RAGE (using Cartesian meshes) introduce perturbations at the mesh scale as the interface moves inward. However, results on the Cartesian meshes compare well with results on FLASH's polar mesh once initial perturbations are introduced.
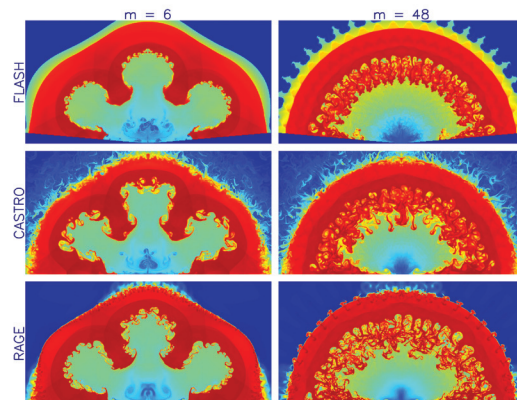
Figure 2 shows the azimuthally-averaged heights of bubbles of light fluid, and depths of spikes of dense fluid, as a function of time in the cylindrical implosions. Again, FLASH's polar mesh introduces no perturbations into an initially unperturbed implosion. Bubble and spike positions agree well across all three codes in the perturbed calculations, although the degree of agreement sometimes varies with time.

Figure 3 shows snapshots of simulated spherical implosions, using the same color scheme as Fig. 1. Because the simulations were performed in 2D RZ coordinates, only half of the sphere was represented, and the interface between light and dense fluids was initialized with even-mode-number perturbations (modes 6 and 48). Bubble and spike amplitudes are again similar across the three codes, but differences in the detailed development of turbulent features are apparent–for example, small-scale features are generally narrowest and sharpest in RAGE.

Ultimately, our goal is to simulate mixing and turbulence in full 3D scenarios. Results from the simulations presented here help us to quantify the spatial resolution required to follow the growth of perturbations of a given initial amplitude and wavelength. The degree of agreement among results on different mesh geometries from codes using different hydrodynamics algorithms establishes requirements for simulating these scenarios in three spatial dimensions. The differences that do appear, in small-scale structure and in specific diagnostics of mixing, could provide specific targets for detailed numerical investigations or future physical experiments.

Fig. 3. Snapshots of simulated spherical implosions.

[1] Dimonte, G. et al., *Phy Fluid* **16,** 1668 (2004).
[2] Arnett, W.D. and C. Meakin, *Astrophys J* **733,** 78 (2011).
[3] Youngs, D.L. and R.J.R. Williams, *Int J Numer Meth Fluid* **56,** 1597 (2008).
[4] Gittings, M. et al., *Comput Sci Discov* **1,** 015005 (2008).
[5] Almgren, A.S. et al., *Astrophys J* **715,** 1221 (2010).
[6] Fryxell, B. et al., *Astrophys J Suppl* **131,** 273 (2000).
[7] Porter, D.H. and P.R. Woodward, "Simulating Compressible Turbulent Flow with PPM," *Implicit Large Eddy Simulation: Computing Turbulent Fluid Dynamics,* Cambridge University Press (2010).

Finding the First Cosmic Explosions

Daniel J. Whalen, T-2, Carnegie Mellon University;
Christopher L. Fryer, CCS-2; Lucille H. Frey, Wesley P. Even,
XTD-6; Catherine Lovekin, T-2; Aimee L. Hungerford, XTD-6;
Christopher J. Fontes, XCP-5; James P. Colgan, T-1

Ten years ago, numerical simulations revealed that the first stars in the universe formed about 200 million years after the Big Bang, ending the cosmic dark ages and beginning the process whereby the universe was gradually transformed from a cold, dark, featureless void into the vast, hot, transparent cosmic web of galaxies we see today. Primordial (or Pop III) stars were also the first great nucleosynthetic engines of the cosmos, creating enormous quantities of the first heavy elements that radically transformed the generations of stars that followed them and making possible the formation of planets and life. Primordial stars are also believed to be one origin for the supermassive black holes that inhabit most massive galaxies today.
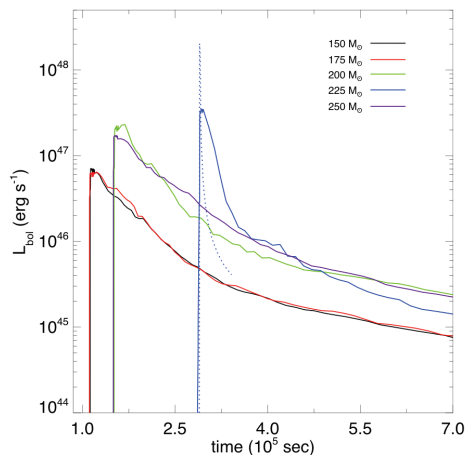
Unfortunately, because primordial stars literally lie at the edge of the observable universe, they remain beyond the reach of both current and planned observatories, so not much is known about their properties or the primeval galaxies they populate. Supercomputer simulations suggest that they were very massive, from 30 to 500 solar masses–if so, stellar evolution models indicate that they died in either core collapse supernova explosions or in far more powerful pair-instability supernovae. The latter are thermonuclear explosions that are up to 100 times more energetic than other supernovae and may be visible at great distances. If next-generation instruments such as the James Webb Telescope (JWST) or the Thirty-Meter Telescope (TMT) glimpse these first cosmic explosions they could place the first constraints on the masses of primitive stars and thus the brightness and colors of the protogalaxies they inhabit.

We have used the LANL radiation hydrodynamics code RAGE (Radiation Adaptive Grid Eulerian) to model the first cosmic explosions in the universe and obtain detailed predictions of their light curves and spectra in order to determine if they will be visible to JWST and the TMT. Hydrodynamical profiles from RAGE were then post processed with the extensive LANL OPLIB database of atomic opacities to compute detailed line-absorption features in our spectra–these constitute the chemical "fingerprint" of the explosion. Both core-collapse and pair-instability explosions were simulated to obtain a comprehensive set of primordial (Pop III) supernovae.

Supernova explosions proceed in several stages. Although they begin deep within the star they cannot be seen until they erupt through its surface, a phenomenon known as shock breakout. As we show in Fig. 1, this extremely hot and energetic shock releases a transient pulse of radiation that is more luminous than the entire Milky Way galaxy. It is mostly composed of X-rays and hard ultraviolet (UV) radiation, and it completely ionizes the wind envelope surrounding the star. Earlier models of supernovae explosions, which treat the fireball as a simple blackbody, predict that the duration of the pulse is comparable to the light-crossing time of the star, since photons simultaneously emitted from its poles and its equator would reach an observer at times that differ by the time it takes light to cross the star, as shown by the dotted blue line in Fig. 1. However, our full radiation-transport models show that the breakout photons remain tightly coupled to gas at the surface of the shock and escape into space over an extended period of time. This creates the dimmer but broader breakout pulse shown in Fig. 1.

The fireball cools as it expands, and there is a gradual shift in the spectrum from short wavelengths to longer ones over time, as we show in Fig. 2. As ejecta from the explosion expands into the intergalactic medium it becomes more diluted, and the photosphere of the explosion (the surface from which photons can escape into space) sinks deeper into the ejecta. As it does, heavier elements deep in the debris imprint absorption lines on the radiation spectrum of the explosion. Pair instability supernovae manufacture vast quantities of radioactive nickel, whose decay produces gamma-rays that downscatter in energy as they slowly diffuse out through the massive ejecta. This heats the ejecta and powers the light curve of the explosion after the initial breakout

Fig. 1. Bolometric luminosities for 150, 175, 200, 225, and 250 solar-mass pair-instability supernovae at shock breakout. The blue dotted line is the approximate lightcurve obtained by treating the shock as a simple blackbody.
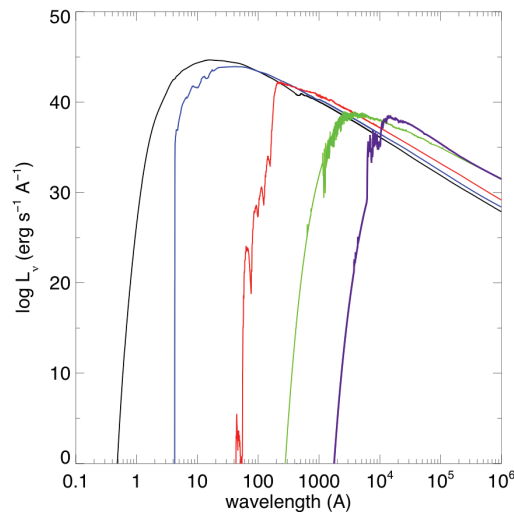
Fig. 2. Evolution of the spectrum of a 250 solar-mass pair-instability supernova. From left to right the times range from 10^4 seconds to three years.

transient. As we show in Fig. 3, the large quantities of nickel synthesized in pair-instability explosions cause them to remain bright for two to three years, in contrast to core-collapse and Type Ia supernovae that fade after a few months. The rebrightening of the explosion that is evident in the more massive explosions at about three months in Fig. 3 occurs when the photosphere of the blast uncovers the hot layer of radioactive nickel deep in the ejecta. The supernova fades after all the nickel has decayed and its photons have diffused out into space.

Radiation from the first supernovae must traverse vast tracts of space over cosmic time to arrive at Earth. Most of the radiation is absorbed by the cosmic web of neutral hydrogen present over most of this epoch, the Lyman-alpha forest. As the universe expanded, the wavelengths of the photons were also stretched to lower energies. Massimo Staivelli, a JWST project leader at the Space Telescope Science Institute, is convolving our Pop III SN spectra with absorption by the Lyman-alpha forest, cosmological redshifting, and JWST NIRCam and NIRSpec filter responses to determine how many of these photons will be captured by JWST. These calculations will reveal out to what redshift (or how far back in time) both JWST and TMT will detect these explosions. If they are discovered at the epoch of first star formation, they will be our first direct probe of the Pop III initial mass function, and hence the character of the first galaxies. Such explosions will also mark the positions of the earliest galaxies in the sky, most of which would otherwise be too dim to be detected. In this and many ways to come, LANL expertise in supernova simulations will open the first direct observational window into the primeval universe.
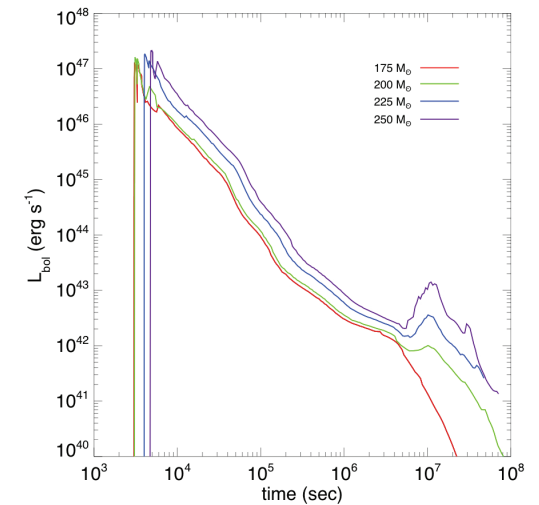
Fig. 3. Bolometric luminosities for 175, 200, 225, and 250 solar-mass pair-instability supernovae out to three years. The resurgence in luminosity at 10^7 seconds in most of the explosions coincides with the descent of the photosphere of the shock into the hot radioactive nickel layer in the ejecta.

References
Frey, L. et al., "The LANL Supernova Light Curve Project," *Astrophys J,* submitted (2012).
Joggerst, C.C. and D.J. Whalen, *Astrophys J* **728,** 129 (2011).
Joggerst, C.C. et al., *Astrophys J* **709**, 11 (2010).
Whalen, D.J. et al., *Astrophys J* **682,** 49 (2008).
Whalen, D.J. et al., "Finding the First Cosmic Explosions I: Pair Instability Supernovae," *Astrophys J,* submitted (2012).
Whalen, D.J. et al., "Detecting the First Cosmic Explosion with JWST," ***Nature***, submitted (2012).

The Birth of the First Quasars

Daniel J. Whalen, T-2, Carnegie Mellon University; Hao Xu, Hui Li, Jarrett L. Johnson, T-2; Christopher L. Fryer, CCS-2; Stirling A. Colgate, T-2

The existence of billion-solar-mass-black holes in massive galaxies less than a billion years after the Big Bang poses one of the greatest challenges to the current paradigm of cosmological structure formation, mostly because it is not understood how they became so massive so early in the universe. Early structure formation is thought to be hierarchical, with small objects at early times evolving into ever more massive ones by accretion and mergers through cosmic time.

It is generally supposed that the supermassive black holes (SMBH) that power the Sloan Digital Sky Survey quasars grew from much smaller seeds at earlier epochs. The origin of SMBHs is of ongoing debate, and three formation channels have been proposed for their seeds: 1) the collapse of primordial stars into 100–300 solar mass black holes 200 million years after the Big Bang, 2) the direct collapse of 10^8 solar mass protogalaxies that have somehow bypassed previous star formation into 10^4–10^6 solar mass black holes about 600 million years after the Big Bang, and 3) more speculative pathways like the relativistic collapse of dense primeval star clusters into 10^4–10^6 solar mass black holes.

The processes by which black holes (BH) form at high redshift and later become supermassive must explain how they become so large just 800 million years after the Big Bang and why their numbers are so small–about one per billion cubic light-years. Although BH-BH mergers likely play a role in SMBH assembly, only accretion can result in the sustained exponential growth required for high-redshift seeds to become supermassive by the time the universe becomes fully ionized at about 1 billion years. Simple Bondi-Hoyle accretion is often invoked to model their growth:

$$\dot{M}_{BH} = \frac{4\pi \rho_\infty G^2 M_{BH}^2}{c_\infty^2 + v_{rel}^2}$$

where ρ_∞ and c_∞ are the density and sound speed of the flow in the vicinity of the BH and v_{rel} is the velocity of the BH relative to the local flow. Because accretion rates depend strongly on mass, seed BHs that form by direct collapse initially grow much faster than primordial (Pop III) BHs, but they form at much lower redshifts and have less time to accrete before reionization is complete.

We are using the Enzo adaptive mesh refinement (AMR) cosmology code to model the collapse of $z \sim 10$–15 protogalaxies into supermassive BH seeds. The process begins when low-mass cosmological halos (spheroidal blobs of dark matter and primordial gas) gravitationally congregate into a small, highly irregular protogalaxy from $z \sim 15$–30. For an SMBH seed to form at the center of this primitive galaxy, none of its constituent halos can previously have hosted star formation because such stars would have evicted all the gas from their halos and the protogalaxy would have been born bereft of gas. The only way this can happen is if the protogalaxy is assembled in close proximity to very strong ultraviolet (UV) sources that sterilize the halos of molecular hydrogen and quench star formation. Such UV backgrounds are thought to be highly unusual in this era, which could explain why SMBHs are rare at $z \sim 6$.

The protogalaxy is hot after it has formed because the kinetic energy its halos once had is converted into heat when they become bound together. Its gas begins to cool by exciting hydrogen (H) and helium (He) line emission, and it pools in the deep gravitational well at the center of the galaxy. Infall rates there become enormous, 0.1–1 solar masses per year, and a stiff hydrostatic core of gas soon forms that rapidly reaches 1000 solar masses or more. At this point it is not known if this object achieves nuclear burning or remains a hot, dead, entropic ball of gas that is later enveloped by an event horizon without ever having become a star. Numerical simulations performed to date to resolve this hydrostatic
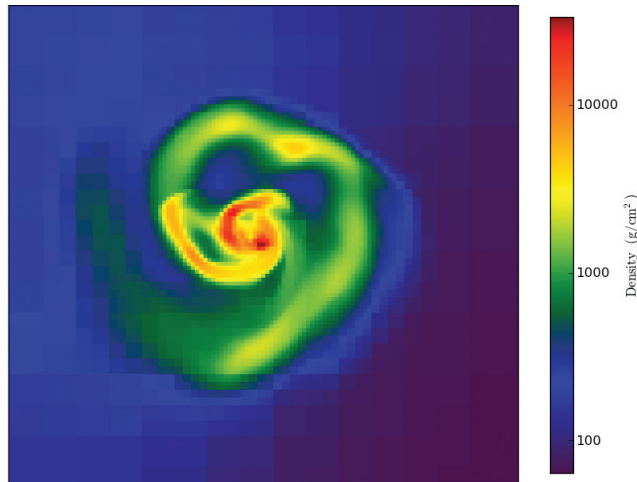
Fig. 1. Formation of an accretion disk around a supermassive star in a collapsing protogalaxy at redshift z = 15.4. The box size is 2000 AU by 2000 AU.

core stall because their time steps become very small and the calculation halts.

We have performed new calculations that prove that if this stiff hydrostatic core of gas does become a massive star, its radiation is always confined close to its surface by the enormous accretion flows raining down on it and cannot propagate out into the galaxy. This result implies that it is not necessary to resolve the hydrostatic seed, just the inflows onto it on larger scales. Retreating from the extreme spatial resolutions previously applied to the seed while preserving the fidelity of the calculation allows us to take much longer time steps in the numerical simulations and evolve the center of the collapsing protogalaxy over many dynamical times. This in turn enables us to determine if infall onto the nascent supermassive star collapses into an accretion disk around it and, later, the conditions under which X-rays from the newborn BH break out into the intergalactic medium after the star dies.

We show Enzo simulations that follow the internal collapse of the protogalaxy in Figs. 1 and 2. They reveal that a massive, atomically cooled accretion disk does form around the central object and rapidly feeds its growth. In parallel with these numerical models, we are pursuing detailed stellar evolution calculations of the fate of the central object with the KEPLER code with Alexander Heger at the University of Minnesota. These simulations will reveal if the object at the center of the disk becomes a star and at what point it becomes a BH. This in turn tells us for how long we must evolve the disk at the heart of the protogalaxy and when we must turn on X-ray emission from the newly formed SMBH seed. These state-of-the-art calculations together with LANL expertise in radiation transport may soon allow us to witness the birth of the first quasars in the universe with supercomputers.
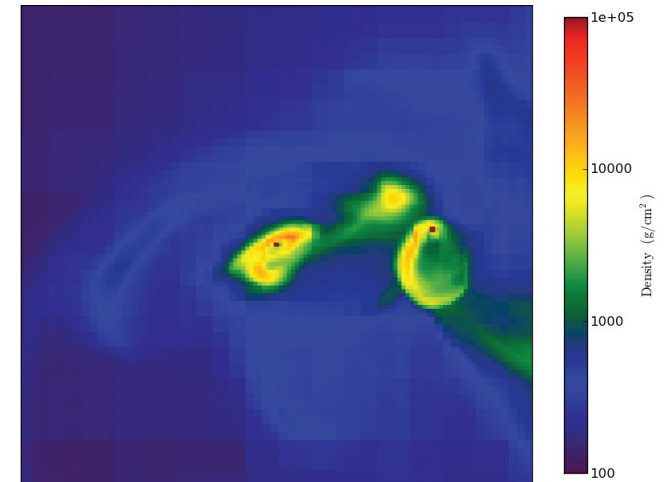
Fig. 2. Fragmentation of the disk into several supermassive Pop III stars.

References
Johnson, J.L. et al., "The Growth of the Stellar Seeds of Supermassive Black Holes," *Astrophys J,* in press (2012).
Whalen, D.J. and C.L. Fryer, "The Formation of Supermassive Black Holes from Low-Mass Pop III Seeds," *Astrophys J*, submitted (2012).
Wise, J.H. et al., *Astrophys J* **682, 7**45 (2008).
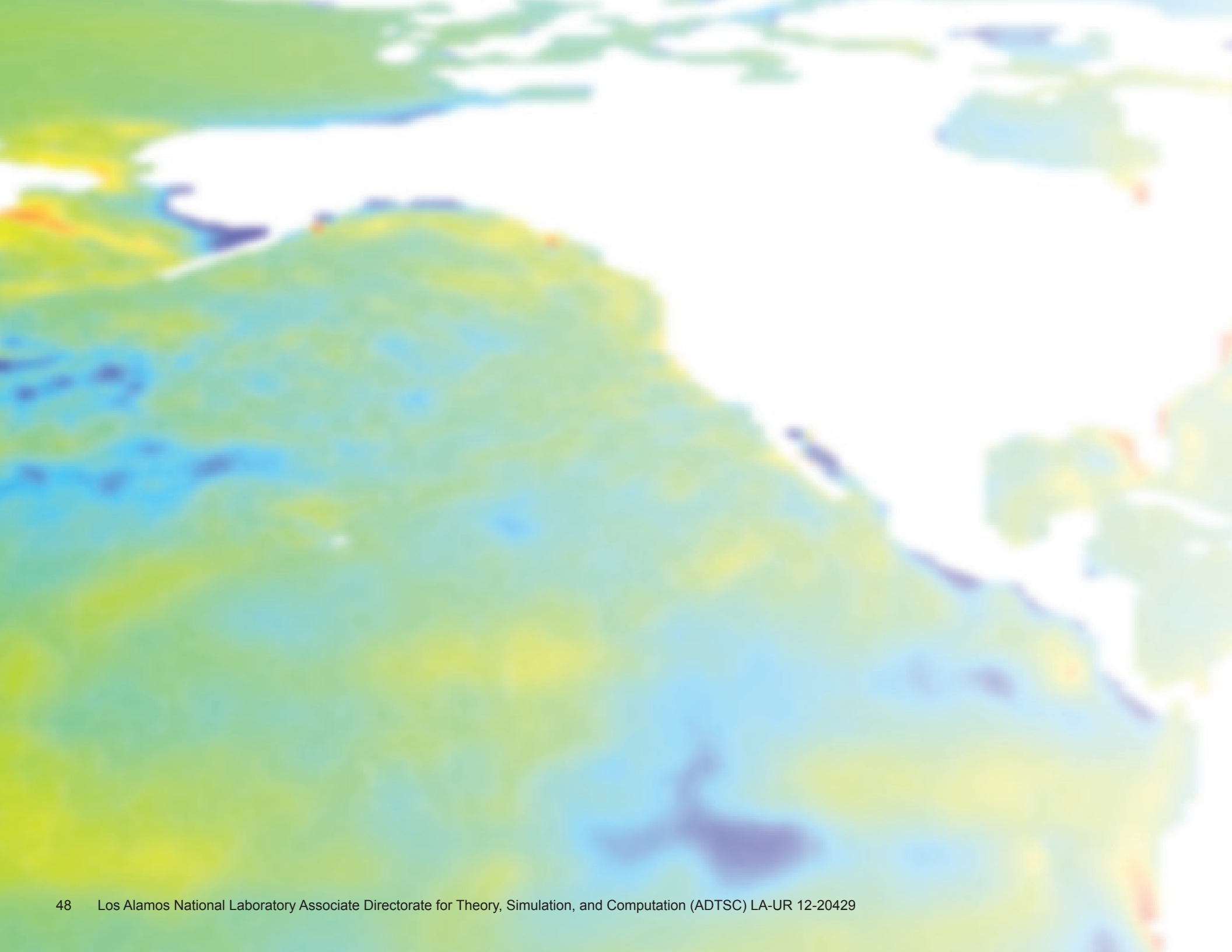
Climate, Atmospheric, and Earth Systems Modeling

LANL is a word-renowned research institution in the area of climate and ocean modeling, stemming from the seminal work on nuclear winter several decades ago that was derived from LANL's nuclear weapons activities.

LANL scientists are responsible for the ocean and sea ice models in the Community Climate System Model (CCSM) and, as such, are also key contributors to Intergovernmental Panel on Climate Change (IPCC) analyses. The articles in this section highlight recent work in modeling the critically important Greenland ice sheet (melt processes and fresh water runoff), implementing enhanced numerical models into LANL codes, creating new advanced numerical methods and models and incorporating them into LANL codes, and quantifying modeling uncertainties through the assimilation of data from a variety of sources. These developments are key in the ongoing quest to understand and quantify our changing climate.

Variable Rates of Acceleration of the Greenland Ice Sheet Due to Penetration of Surface Meltwater

Matthew J. Hoffman, T-3; Ginny A. Catania, University of
Texas; Thomas A. Neumann, NASA Goddard Space Flight
Center; Lauren C. Andrews, University of Texas;
Julie A. Rumrill, Southern Connecticut State University

Summer melt in the ablation zone of the Greenland Ice Sheet penetrates through ice 1 km thick, lubricating the bed and causing a summer speedup. Should continued warming reinforce this process, faster flow could cause drawdown of the ice sheet surface to lower, warmer elevations, providing a positive feedback. We compare subdaily ice velocity and uplift derived from nine global positioning system stations in the upper ablation zone in west Greenland to surface melt and supraglacial lake drainage during the summer of 2007. A speedup of over 30%, lasting approximately 40 days, occurred across the region. Sliding appears to be driven by diurnal and episodic pulses of meltwater delivery to the bed. The impact on velocity from surface melt decayed as the season progressed, indicating evolving efficiency of the subglacial hydrologic system. Thus, increases in seasonally averaged summer melting may not guarantee faster seasonal ice flow.
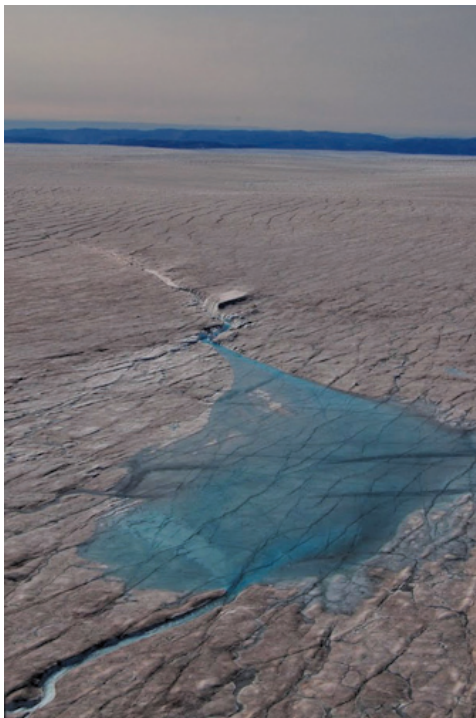
Ice mass loss from the Greenland Ice Sheet is raising global sea level by about 0.8 mm per year and is accelerating [1]. Basal lubrication by surface meltwater penetrating the 1000 m-thick ice sheet (Fig. 1) generates speedups during summer of 50–200% for the regions of the ablation zone experiencing sheet flow [2,3]. This seasonal acceleration has led to uncertainty in the dynamic sensitivity of the Greenland Ice Sheet mass balance to increased atmospheric temperatures. If surface melting translates directly to increased sliding, rising temperatures could generate a positive feedback to mass loss as faster flow causes the ice sheet to lower into warmer elevations [2,4]. On the other hand, if the subglacial hydrologic system of the ice sheet adapts to variable meltwater input, as observed on mountain glaciers, increased melt could generate a limited, or even decelerating, effect on seasonally averaged sliding and long-term dynamic response to warming climate [5,6]. Additionally, rapid drainage to the bed of water in lakes that form on the ice sheet surface can temporarily increase ice velocity by over an order of magnitude [7].

A better understanding of these processes is necessary for the next generation of ice sheet models to accurately predict the evolution of the Greenland Ice Sheet over the next century and its contribution to sea-level rise. In a modeling study where increases in ice flux were assumed to be proportional to increases in surface melt, the Greenland Ice Sheet was projected to contribute to sea level rise at rates approximately 20% higher than when the effect is not included [4]. However, more recent work suggests the relationship between surface melting and sliding is not that simple [5,6] and, if so, such an approach may vastly overestimate mass loss from the ice sheet.

We investigated the seasonal and regional evolution of the melt-driven summer speedup of the Greenland Ice Sheet using velocity measured at nine global positioning system (GPS) stations across a study area spanning approximately 50 km of the western Greenland Ice Sheet (Fig. 2). Ice sheet velocity was compared to modeled surface melt and an inventory of lake drainages.

In the ablation zone of the ice sheet, there was a speedup of close to 40% during the summer, followed by a slight slowdown in late summer (Fig. 3), while acceleration was minimal close to the equilibrium line (stations 607–807). The regional speedup began within a few weeks of melt onset and persisted as long as air temperatures remained above freezing. Melt is well correlated with acceleration, but the magnitude of acceleration relative to melt decreases dramatically as the season progresses, indicating an evolution of the ability of the subglacial hydrologic system to accommodate the input of surface meltwater [6].

Subdaily velocity variations do not indicate a sustained season-long speedup, but rather a series of diurnal speedups, with nighttime velocities dropping near to winter values (Fig. 3). In fact, minimum daily velocity is somewhat higher than winter velocity initially, after the

Fig. 1. Meltwater lake draining water to the bed of the Greenland Ice Sheet.
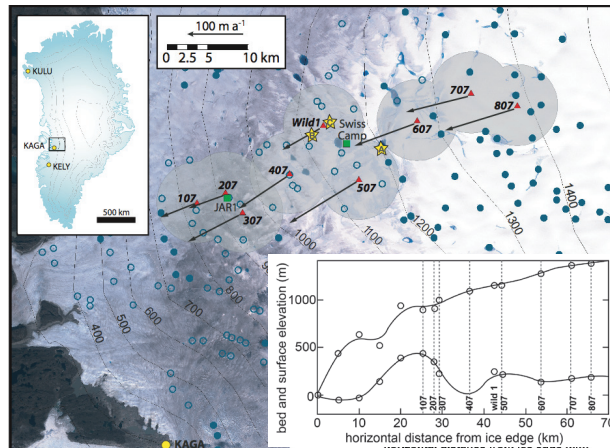
Fig. 2. Locations of GPS stations (red triangles), weather stations (green squares), and lakes (blue circles) shown on a LandSat image. Open/filled circles indicate lakes that drained/filled during the June 22–July 8, 2007 time period. Inset shows ice sheet surface and bed elevation along the GPS station flowline.

Fig. 3. Six-hour mean horizontal velocity (dark blue, green, and red lines) and hourly air temperature (gray) for GPS stations. Velocity has been normalized with respect to winter velocity.
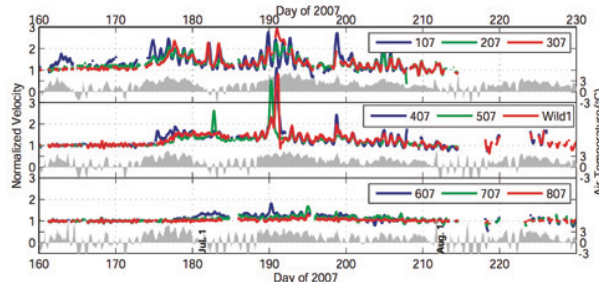
seasonal speedup begins, but gradually lowers as the summer progresses (Fig. 4). However, the amplitude of daily speedup above the minimum velocity correlates strongly with daily melt production (Fig. 4). Thus, daily pulses of meltwater appear to drive daily acceleration, despite a clear seasonal evolution of the subglacial hydrologic system.

The dynamic response to meltwater input is similar to that observed on mountain glaciers. However, the ice in our study area is much thicker–about 1 km thick–which should cause creep closure within hours of efficient subglacial channels that develop beneath the ice to convey water down-glacier. The fact that the subglacial hydrologic system evolves to increasingly greater efficiency as the summer progresses suggests that water pressures remain high enough to offset much of the overburden pressure of the ice, effectively making the ice sheet behave like thinner ice.

Over the summer we identified three lake drainage events in the region that locally increase daily velocity by two to three times winter velocity and result in significant surface uplift. However, less than half of the lakes greater than 0.5 km in diameter generated detectable acceleration; lake drainage can explain at most 5% of the total summer motion at any station, and it is therefore not an important component of the seasonal speedup.

Our observations support the ideas that rising air temperatures in Greenland may not translate directly into increased sliding at the seasonal scale [5,6] and that episodic pulses of water are key for generating enhanced sliding [5]. Daily pulses of meltwater appear to cause variations in daily sliding that are broadly consistent over space and time in our study area, but these daily increases in

velocity are superimposed on a nighttime velocity that generally decreases over the season, presumably as the subglacial hydrologic system becomes more efficient. Thus, accurately predicting future dynamics of the ice sheet in response to changes in surface melt may require knowledge of both melt generation and the state of the subglacial hydrologic system at high temporal resolution. Ongoing work involves introducing a realistic subglacial hydrology model that captures these effects in the ice sheet model used by the Climate, Ocean and Sea Ice Modeling (COSIM) project for modeling predictions of sea level rise from the Greenland Ice Sheet.

Fig. 4. Minimum daily velocity (black) and difference between maximum and minimum daily velocity (dark gray) for stations a) 107, and b) 407. Modeled daily melt is shown with light gray shading.

References
Hoffman, M.J., G.A. Catania, T.A. Neumann, L.C. Andrews, and J.A. Rumrill, "Links between Acceleration, Melting, and Supraglacial Lake Drainage of the Western Greenland Ice Sheet," *J Geophys Res*, **116,** 16 (2011).

[1] Velicogna, I., *Geophysic Res Lett* **36**, L19503 (2009).
[2] Zwally, J. et al., *Science* **297,** 218 (2002).
[3] Joughin, I. et al., *Science* **320,** 781 (2008).
[4] Parizek, B.R., and R.B. Alley, *Quaternary Sci Rev* **23,** 1013 (2004).
[5] Schoof, C., *Nature* **468,** 803 (2010).
[6] Bartholomew, I. et al., *Nat Geosci* **3,** 408 (2010).
[7] Das, S. et al., *Science* **320**, 778 (2008).

Parallel Algorithm for Spherical Centroidal Voronoi Tessellations

Douglas W. Jacobsen, T-3; Max Gunzburger, Florida State University; Todd D. Ringler, T-3; John Burkardt, Janet Peterson, Florida State University

Spherical centroidal Voronoi tessellations (SCVT) are used in many applications in a variety of fields, one being climate modeling. They are a natural choice for spatial discretizations on the Earth, or any spherical surface. The climate modeling community, which has started to make use of SCVTs, is beginning to focus on exascale computing for large-scale climate simulations. As the data size increases, the efficiency of the grid generator becomes extremely important. Current high resolution simulations on the earth call for a spatial resolution of about 15 km. In terms of an SCVT this corresponds to a quasi-uniform SCVT with roughly 2 million Voronoi cells. Computing this grid serially is very expensive and can take on the order of weeks to converge sufficiently for the needs of climate modelers. Outlined here is a new algorithm that utilizes existing computational geometry tools such as conformal mapping techniques, planar triangulation algorithms, and basic domain decomposition, to compute SCVTs in parallel, thus reducing the overall time to convergence. This new algorithm shows speedup on the order of 4000 when using 42 processors over STRIPACK in computing a triangulation used for generating an SCVT.

Recent developments in ocean and atmospheric modeling allow for simulation on Spherical Centroidal Voronoi Tessellations (SCVT) [1]. However, SCVTs cannot simply be prescribed, they have to be generated. In order to generate a SCVT, algorithms make use of their dual mesh, Delaunay Triangulations. Typically SCVTs are generated using what is known as Lloyd's algorithm, which can be either deterministic of probabilistic [2]. A simple version of this algorithm would be as follows:

1) Start with a point set
2) Determine the Delaunay Triangulation of the point set
3) Determine the Voronoi Diagram of the point set
4) Determine the center of mass for each Voronoi Region
5) Replace the point set with the centers of mass
6) Iterate until converged

A Voronoi diagram defines regions where all points contained inside of a region are closer to the region's center than they are to any other point in the generating set. A special type of Voronoi diagram is known as a centroidal Voronoi tessellation, which occurs when the center of mass of a Voronoi region is coincident with the region's generating point. The spherical variant of a centroidal Voronoi tessellation is the same as a planar version, but on the surface of a sphere. Current software for determining a spherical Voronoi tessellation, such as STRIPACK, are limited in the size of meshes they can compute as well as the speed at which they can determine the tessellation.

In order to generate high resolution meshes for ocean and atmospheric simulations, a new algorithm was developed to compute SCVTs in parallel. In order to generate a SCVT in parallel two things are done with the point set. First, domain decomposition needs to be used. In order to simplify this process, a SCVT that is coarser than the target SCVT is used to decompose the surface of the sphere into regions. After this is done, tangent planes are defined at each of the region centers. Using these tangent planes, points within a region are stereo-graphically projected (Fig. 1) into the plane and the Delaunay triangulation is computed in a plane. This triangulation can then be mapped onto the surface of the sphere with only small modifications, including the removal of triangles whose circumcircles extend outside of the region's radius. This triangulation is valid because stereographic projections preserve circularity, keeping the insides and outsides of circles through the projection. The only restrictions in order for a triangle to be Delaunay involve its circumcircle being empty, and having three or more points on its perimeter.

Fig. 1. Cross-section of a stereographic projection from a sphere into a plane.
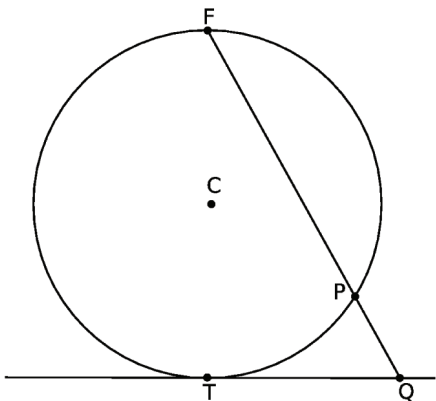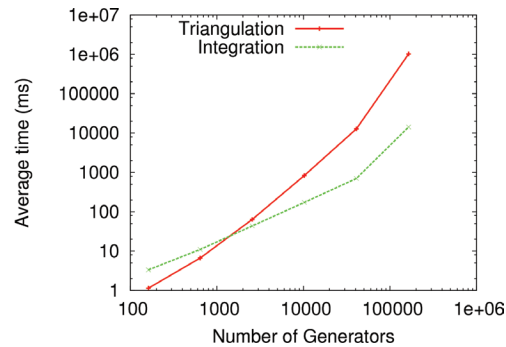
Fig. 2. Timing for different portions of a STRIPACK-based SCVT generator.

Using these techniques the new algorithm for generating an SCVT in parallel is as follows:

1) Start with a point set
2) Decompose point set into regions
3) Stereographically project point sets into tangent planes
4) Triangulate points in tangent planes
5) Map triangulations onto sphere
6) Determine centers of mass of Voronoi regions
7) Replace point set with centers of mass
8) Iterated until converged

During the decompose step, a point might exist in more than one region. This causes each region to overlap with its neighbors and adds a small buffer region that ensures the interior of the region contains the exact Delaunay triangulation and Voronoi diagram for the points that need to be updated. A naive approach is to define a radius for each region, and sort each point using a dot product with the center. If the dot product results in a distance that is smaller than the region radius that point will be part of that region's computational domain. This overlap is the reason that certain triangles need to be removed from the set. If a triangle's circumcircle extends outside of its region radius, points outside of the current set might be contained inside of the circumcircle, making it no longer Delaunay. After points are triangulated, a region only updates points contained within its Voronoi region.

Fig. 3. Timing for different portions of a modified SCVT generator, using two processors.
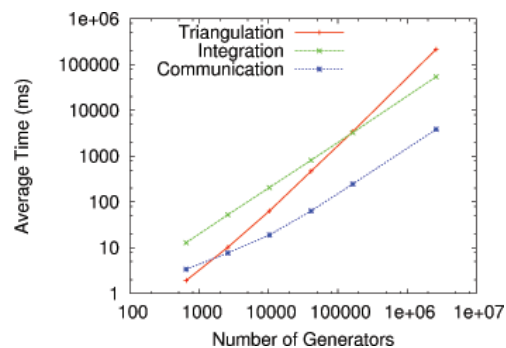
This modified version of Lloyd's algorithm can now be used to generate SCVTs in parallel using distributed memory systems. An implementation of this algorithm is compared with an implementation of Lloyd's algorithm using STRIPACK for serial triangulations. When using the serial version, the cost of computing a triangulation increases non-linearly as the problem size increases (Fig. 2). Whereas the parallel version provides linear increases in both the cost of triangulation and the cost of computing the centers of mass when using only two processors (Fig. 3). Adding more processors increases the speedup when compared with the STRIPACK version to something on the order of O(4000), and this
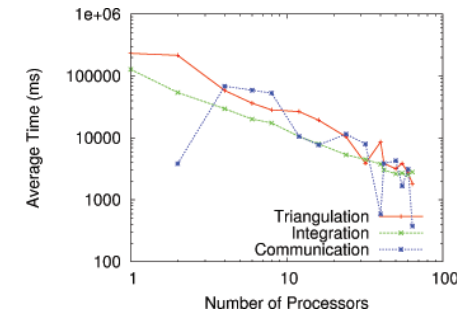
Fig. 4. Scaling for the modified SCVT generator using 2621442 generators.

algorithm shows reasonable scaling with when generating a 2.6-million-point mesh (Fig. 4).

One issue that arises when exploring the generation of variable resolution meshes with this algorithm amounts to poor load balancing. This problem is fixed by first sorting points into Voronoi regions for the coarse SCVT. After this initial sorting is complete, a regions point set consists of the union of the points contained in that region and all of that neighboring regions.

Using this new algorithm, high-resolution meshes can now be generated in parallel on the surface of the sphere for use in ocean and atmosphere models. In addition to providing the capabilities of high resolution mesh generation, lower-resolution meshes can be generated in significantly less time than previous techniques provided. This method can also be ported to be used in a plane by removing the stereographic projection portion, and it may be possible to also extend it to allow parallel triangulations of an entire 3D sphere using some other techniques.

[1] Ringler, T.D. et al., *Mon Weather Rev* **139,** 3348 (2011).
[2] Lloyd, S. *IEEE Trans Inform Theor* **28(2)**, 129 (1982).

Spatial Freshwater Runoff Distribution from Greenland, 1960–2010

Sebastian H. Mernild, CCS-2

Fig 1. (a) Greenland with topography (500-m contour interval) and the location of the coastal and GrIS meteorological tower stations (red dots). HydroFlow simulations were performed for Greenland using a 5-km grid increment and daily time step from September 1959–December 2010, forced with observed atmospheric data from 56 meteorological stations located in coastal areas and on the ice sheet. The ice sheet is shaded from gray to white by elevations, with the area outside the ice sheet in black. (b) Simulated individual Greenland drainage basins (represented by multicolors). Also, a specific region is illustrated (see bold square). (c) A close-up example of the individual drainage basins and flow network for the Helheim Glacier region, at the innermost part of the Sermilik Fjord, SE Greenland.
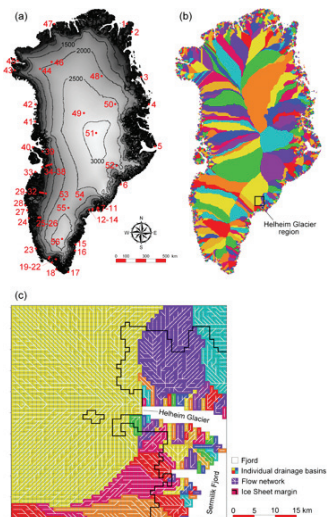
HydroFlow, a gridded runoff routing model, was developed to simulate the linkages between runoff production from land-based snow and ice-melt processes and the associated freshwater fluxes to downstream areas and surrounding oceans. HydroFlow was specifically designed to account for ice sheet, glacier, and snow-free and snow-covered land applications. As part of its discharge simulations, HydroFlow creates a flow network that links the individual grid cells that comprise the simulation domain. Runoff magnitudes, the spatial patterns from individual Greenland catchments, and their changes through time (1960–2010) were simulated in an effort to understand runoff variations and trends to adjacent seas. Total Greenland runoff to the surrounding oceans increased 30%, averaging 481±85 km³ per year. Averaged over the period, 69% of the runoff to the surrounding seas originated from the Greenland Ice Sheet and 31% came from outside the ice sheet from melting glaciers and ice caps. Regionally, runoff was greater from western than eastern Greenland.

Long-term temperature observations show warming trends of variable strength throughout the Arctic, and ample evidence indicates that the Arctic hydrological cycle, including that for Greenland and the Greenland Ice Sheet, is changing [1-3]. Since the early 1990s, the increase in ice sheet surface runoff has followed atmospheric warming, explaining half of the recent mass loss of the Greenland Ice Sheet [4] —a mass loss that by 2100 may contribute up to a 54-cm sea-level equivalent (SLE) [5] because model simulations of future scenarios point to higher temperatures [2,6]. In addition to the sea level contribution, terrestrial runoff from Greenland is important for ocean density and thermohaline circulation [7], specifically to the Atlantic Meridional Overturning Circulation (AMOC) and its impact on the climate system [8].

HydroFlow is a spatially distributed model that divides the simulation domain into individual drainage basins, linking each grid cell within each drainage basin via an eight-compass-direction waterflow routing network. As part of the flow network generation, only a single flow outlet into the ocean is allowed for each individual watershed. The waterflow is transported through the gridded routing network via linear reservoirs, and conservation of mass principles between inflow, storage change, transit times, and outflow from each cell in the routing network must be defined to simulate the catchment runoff and generate discharge hydrographs for the routing grid cells. For runoff routing, HydroFlow assumes the existence of different transport mechanisms within each individual model grid cell: 1) a slow-response runoff system, representing the time it takes for any available snow and ice melt, including liquid precipitation, to be transported within a model grid cell to the fast-response reservoir; and 2) a fast-response

system representing flow processes such as those represented by water transport over, inside, and below the ice that moves water down-network. As part of the modeling system, locally generated runoff from snow-covered ice, snow-free ice, snow-covered land, and snow-free land are all associated with different residence times that evolve with time as the snow and ice melt [9].

HydroFlow divided the ice sheet into approximately 400 individual drainage basins, and Greenland into approximately 3,150 individual basins, where the drainage basins range in area from 50 km² to 154,800 km² (Fig. 1), with 85% of the drainage basins equal to or less than 250 km²—these relatively small basins cover 10% of the total Greenland area, and are mainly located in the land area between the ice sheet and the oceans.

Time series (1960–2010) of Greenland Ice Sheet surface hydrological conditions of net precipitation (precipitation minus evaporation and sublimation), surface runoff, and surface mass balance are shown in Fig. 2. The average 1960–2010 simulated ice sheet net precipitation was 489±53 km³ per year, varying on a decadal scale from 456±46 km³ per year in 1960–1969 to 516±38 km³ per year in 2000–2010. The ice sheet average runoff was 333±75 km³ per year, varying from 261±75 km³ per year in 1970–1979 to 429±57 km³ per year in 2000–2010. For the period 1960–2010, the Greenland Ice Sheet net precipitation and surface runoff rose significantly, leading to a significantly enhanced surface mass loss. Even with an enhanced surface mass loss, the average annual surface mass balance was 156±82 km³ per year (1960–2010), varying from 220±86 km³ per year in 1970–1979 to 86±72 km³ per year in 2000–2010. These surface conditions closely follow air temperature
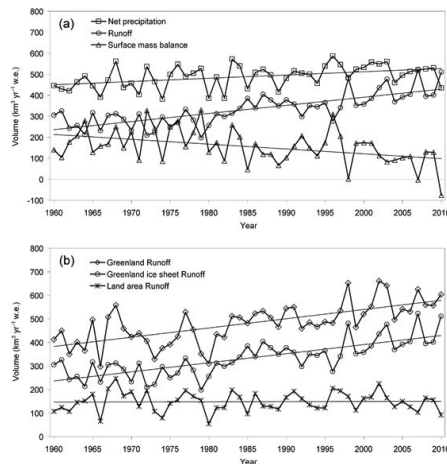
Fig. 2. (a) Greenland Ice Sheet simulated net precipitation, surface runoff, and surface mass balance time series for 1960–2010. (b) Simulated surface ice sheet runoff, land strip area (area outside the ice sheet) runoff, and Greenland runoff time series for 1960–2010.

Fig. 3. Simulated Greenland spatial surface runoff for the decade with the lowest (1970–1979) and highest (2000–2010) runoff, and the 50-year mean (1960–2010).
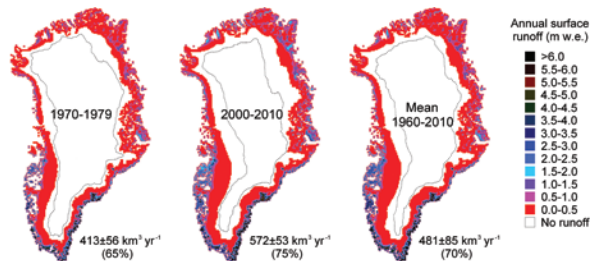
fluctuations, indicating that surface mass loss increased as climate warmed with no suggestion of deceleration.

Time series of Greenland runoff (1960–2010) and individual runoff contributions from the ice sheet and from the land area–including thousands of glaciers and ice caps–located between the ice sheet and the surrounding oceans are shown in Fig. 2b. The 1960–2010 average simulated Greenland runoff was 481±85 km^3 per year, indicating that 69% of the runoff to the surrounding seas originated from the ice sheet and 31% from the land area. For the land area, the trend in runoff was constant, and the average runoff was 148±41 km^3 per year.

For Greenland, the spatial distribution of surface runoff is illustrated for the decade with the lowest (1970–1979) and highest (2000–2010) runoff, and for 1960–2010 (Fig. 3). Generally, relatively high average surface runoff values were simulated for the southwest and southeast regions of Greenland, and sporadic high values were simulated in the north region with maximum values of 4–6 m w.e. per year. Elsewhere runoff was less, with lowest values in the northeast and northwest regions of less than 0.5 m w.e. per year. This regional pattern in surface runoff can be largely explained by the spatial distribution of precipitation, since snowfall (end-of-winter accumulation) and surface runoff are negatively correlated through surface albedo, snow depth, and snow characteristics (e.g., snow cold content) [1,10,11].

In Fig. 4, the spatial runoff distribution from Greenland to the adjacent seas is illustrated. The 1960–2010 average discharge for these drainage catchments varied from <0.01 km^3 per year to 10.1 km^3 per year. The spatial variability in catchment runoff to the surrounding seas varied according to catchment size, ice sheet and glacier elevation range, and ice sheet and glacier areal coverage within each individual catchment. For approximately half of the catchments (colored radial bars), runoff amounted to less than 1.0 km^3 per year (1960–2010), collectively contributing 15% of the 481-km^3-per-year total Greenland runoff. In contrast, 15% of the catchments–catchments having a relatively large ice sheet and/or glacier areal coverage–had a mean

annual runoff greater than 2.5 km^3 per year and contributed 40% of the Greenland runoff to the adjacent seas. Greenland was further divided into eight 45-degree regions. Regionally, the average Greenland 1960–2010 simulated runoff to the adjacent seas was greatest from sectors S (88 km^3) and SW (82 km^3), and lowest from sectors E (45 km^3) and SE (49 km^3). The length of the discharge season at regional scales was highest in the southern sectors (averaging approximately 4–6 months) and lowest in the northern sectors (averaging approximately 2–3 months).

Fig. 4. Spatial distribution of simulated runoff from Greenland's individual drainage basins (each radial colored bar represents the accumulated runoff of ten catchments located side by side [in total there are 316 radial bars]; this was done to simplify the presentation of spatial runoff trends, since 85% of all catchments are equal or below 250 km^2), and from the eight sectors (N, NE, E, etc.), to the adjacent seas: mean annual Greenland runoff for 1960–2010. The numbers in parentheses indicate the length of the discharge season for each region.
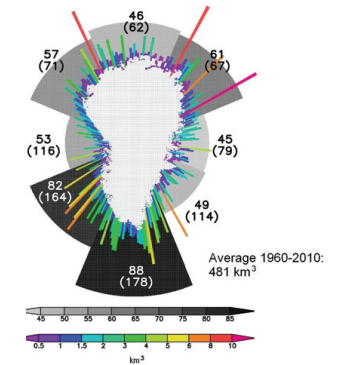
[1] Ettema, J. et al., *Geophys Res Lett* **36**, L12501; doi:10.1029/GL038220 (2009).
[2] Mernild, S.H. et al., *J Hydrometeorology* **11**, 3 (2010).
[3] Mernild, S.H. et al., *J Climate* **24**, 609; doi: 10.1175/2010 JCLI3560.1 (2011).
[4] Van den Broeke, M.R. et al., *Science* **326**, 984 (2009).
[5] Pfeffer, W.T. et al., *Science* **321**, 1340 (2008).
[6] Solomon, S.D. et al., Eds., *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change,* Cambridge University Press, Cambridge, UK and New York, NY, USA (2007).
[7] Rahmstorf, S. et al., *Geophys Res Lett* **32**, L23605; doi:10.1029/2005GL023655 (2005).
[8] Bryden, H.L. et al., *Nature* **438**, 655; doi:10.1038/nature04385 (2005).
[9] Mernild, S.H. and G.E. Liston, *J Climate,* accepted (2012).
[10] Hanna, E. et al., *J Climate* **21**, 331 (2008).
[11] Mernild, S.H. et al., *Hydrolog Process* **23**, 2470; doi: 10.1002/hyp.7354 (2009).

Improving Climate Predictions through Ocean-data Assimilation

Balasubramanya T. Nadiga, William R. Casper, CCS-2;
Philip W. Jones, T-3

Real-world observations have to continually inform the increasingly sophisticated computational models of the climate system in order for the model trajectories to be related to the actual trajectory of the climate system (now routine in numerical weather prediction). A prerequisite for the successful combination of these two streams of information is a proper quantification of their respective uncertainties. An application of this procedure over the period for which observations are available is expected to improve climate predictions as well as help identify system sensitivities and modeling deficiencies.

Fig. 1. Area weighted average of the sea-surface temperature (SST) in the North Atlantic from 0 to 70 N using monthly Kaplan SST, and its 10-year running average–the AMO index [1]. The AMO affects air temperatures and precipitation over much of the Northern hemisphere, is associated with changes in the frequencies of North American droughts and severe Atlantic hurricanes, and renders the identification of anthropogenic climate change signals difficult. While links to oceanic thermohaline circulation have been hypothesized, the predictability of AMO is an area of active research. Nevertheless, proper initialization of the phase and amplitude of such large-scale features through the process of data assimilation is necessary to improve climate predictions.
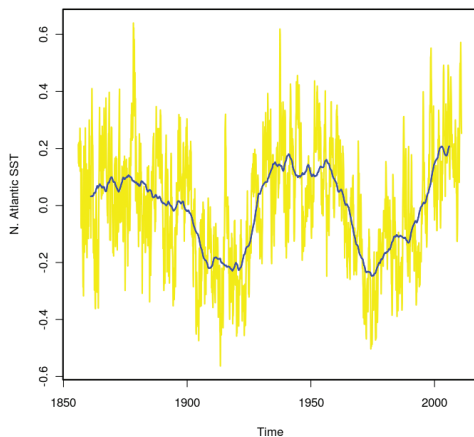
On the interannual and longer time scales, ocean circulation mediates variability of the climate system because of its immense dynamical and thermodynamical inertia. Given such control of climate variability by ocean circulation, it is necessary that ocean models be initialized with conditions that represent as accurately as possible the actual ocean state, including the phase and amplitude of possible low-frequency oscillations (e.g., the Atlantic Multidecadal oscillation [AMO] in Fig. 1) in order to improve reliability of climate predictions.

However, global ocean circulation is a strongly coupled multiscale system with myriad physical processes occurring over a wide range of spatial and temporal scales that interact to give rise to complicated low-frequency variability. While there have been enormous advances in the understanding and modeling of global ocean circulation over the past half a century, an inescapable aspect of ocean circulation as regards prediction is the chaotic nature of the underlying dynamics. Consequently, the state of the modeled ocean circulation has to be continually reinitialized in order for the evolution of the modeled system to be related to that of the actual system. The periodic reinitialization is effected by the process of data assimilation wherein model forecasts and real-world observations are combined using their respective uncertainties.

We have recently developed an ensemble-based data assimilation system for the LANL ocean model POP (Parallel Ocean Program) using the Data Assimilation Research Testbed [2] of the National Center for Atmospheric Research (NCAR). In this approach, uncertainty of model-based prediction is quantified using an ensemble of model forecasts (highly computationally intensive). Notwithstanding the fact that such a method has the highly attractive feature of being able to better estimate uncertainty due to dynamic instabilities associated with the evolving flow, a frequent problem that plagues ensemble-based ocean data assimilation is that the RMS error grows faster than the ensemble spread.

As a consequence, the ensemble spread collapses all too soon leaving little room for the analysis step to make appropriate use of the available observations. In other words, the poor or collapsed ensemble spread is interpreted by the filter as high certainty in the model forecast, forcing it to neglect observations–in turn leading to increased root-mean-square (RMS) error–and an eventual failure of the assimilation. We have developed schemes to remedy such pathological divergence of model trajectories from observations by hybridizing dynamic ensemble-based estimates of uncertainty (which account for uncertainty due to those in initial conditions and those due to dynamic instabilities) with static estimates of uncertainty that account, in part, for model errors [3]. Successful assimilation of ocean observations was possible only with these improvements; the results of assimilating World Ocean Database observations over a 16-month period starting in 1990 in POP are shown in Figs. 2 and 3. The resulting improvement in prediction skill is shown in Fig. 4. In future work, we will use more recent ARGO float-based observations and develop hybrid schemes to enable assimilation of ocean observations in higher-resolution POP simulations than the one-degree resolution that we presently use.
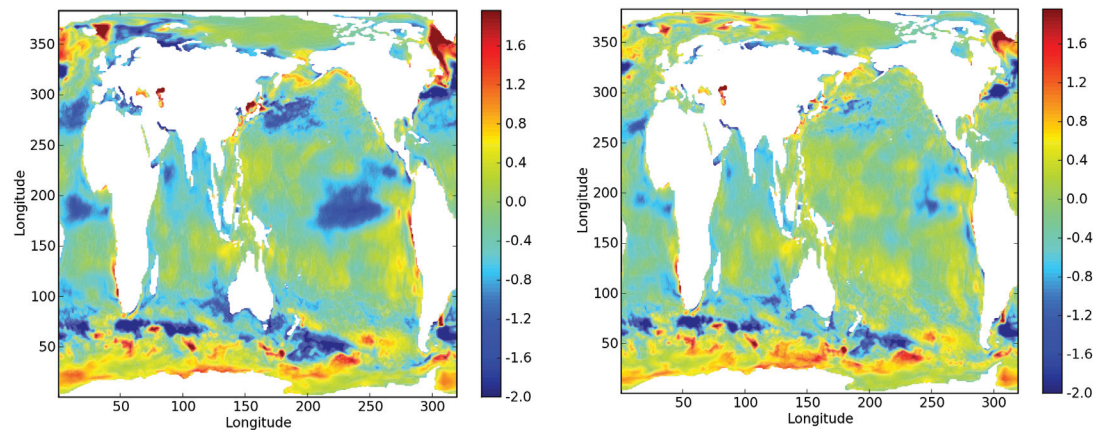
Fig. 2. Left: SST error of control ensemble run without assimilation. Error is with respect to NOAA OI SST v2 and is averaged over the 16-month period of the experiment. Note, cold bias in tropics and midlatitudes and warm bias at high latitudes and upwelling regions. Right: SST errors are reduced on successfully assimilating World Ocean Database observations. Note that NOAA OI SST v2 is not assimilated.

Fig. 4. Successful data assimilation improves skill over up to a year. Skill exhibits a fast decay over the first six months, followed by a slower decline.

Fig. 3. Area-weighted mean absolute error of the monthly-averaged SST anomaly for January 1990 through April 1991. Left: A sophisticated statistical inferencing procedure fails to effect improvements over control in this assimilation experiment. Right: Hybridizing dynamic ensemble-based estimates of uncertainty with a static estimate leads to successful assimilation and significant reduction in error over control.

[1] Enfield, D.B. et al., *Geophys Res Lett* **28**, 2077 (2001).
[2] Anderson, J.L et al., *Bull Am Meteorol Soc* **90**, 1283 (2009).
[3] Nadiga, B.T. W.R. Capsper, and P.W. Jones, "*Ocean Data Assimilation: An Ensemble Data Assimilation System for the Parallel Ocean Program,*" Los Alamos National Lab. Technical Report LA-UR-12-10448

Physics-based Preconditioners for Ocean Simulation

Christopher K. Newman, Dana A. Knoll, T-3

Physics-based preconditioning is a highly successful approach for multiple-time-scale problems where an accurate simulation is desired on the dynamical time scale. In our research we are developing physics-based preconditioners for ocean simulation based on barotropic-baroclinic splitting. Our approach is a fully implicit, fully coupled time integration of the momentum and continuity equations of ocean dynamics, thus reducing the error and increasing the stability of traditional operator splitting. The nonlinear system is solved via preconditioned Jacobian-free Newton-Krylov, where we reformulate traditional barotropic-baroclinic splitting as a preconditioner. Thus the desired solution is time step converged with time steps on the order of the dynamical time scale, a crucial feature as we scale to exascale simulation.

Fig. 1. Timestep size versus CPU time for explicit Runge-Kutta (RK4), unpreconditioned implicit Euler (JFNK) and preconditioned implicit Euler (JFNK-CG-ML). Implicit methods afford a stable integration with much larger time steps than explicit methods; preconditioned implicit Euler utilizes time steps on an order of 30 times the barotropic time scale. However, the key to efficient, scalable implicit methods is effective preconditioning; implicit Euler preconditioned with the barotropic system scales on the same order as Runge-Kutta, while unpreconditioned implicit Euler does not scale beyond the barotropic timescale.
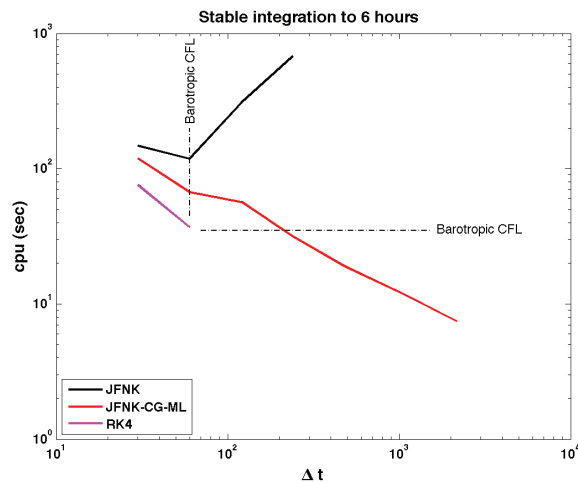
Due to numerical stability issues, low-order accuracy and inherent step-size restrictions associated with traditional explicit, split-explicit, and semi-implicit time integration methods, implicit time integration will be required for high-resolution ocean modeling. We are developing physics-based preconditioners for implicit ocean simulation that allow enhanced accuracy, stability, and the ability to time integrate with time steps much larger than explicit or semi-implicit methods currently in practice. Our implicit time integration is implemented with Jacobian-free Newton-Krylov (JFNK) as a nonlinear solver [1]. The JFNK framework allows tighter coupling of the physics, thus reducing errors and increasing stability inherent to operator splitting. In addition, higher-order implicit time integration schemes are easily implemented in the JFNK framework. The key to effective implementation of JFNK is effective preconditioning. As time-step sizes in explicit and split-explicit methods are dictated by the barotropic time scale, we have implemented a barotropic solver as a physics-based preconditioner, based on implicit time integration of the barotropic system over the baroclinic time interval, which allows efficient implicit time integration on the order of the baroclinic time scale.

Following the standard prescription for physics-based preconditioning, we reformulate the current ocean simulation semi-implicit solver [2] as a preconditioner. The ocean model is decoupled into baroclinic and barotropic systems. The barotropic system is a 2D hyperbolic system (obtained by vertical integration of the full system) that isolates the stiff external gravity wave and can easily be decoupled to a parabolic system for a scalar variable that can be integrated implicitly over the baroclinic time step. This implicit integration necessitates solving a linear system involving an elliptic operator, which can be implemented in a scalable way using multigrid solver technology. As a result, application of the preconditioner is reduced to a scalable elliptic solve for a 2D scalar variable, rather than solution of a full 3D vector system. This approach allows a split-implicit solver (as found in Parallel Ocean Program, POP), to be used as a solver or preconditioner.

The JFNK framework is implemented in a z-level ocean model prototype via an interface to the nonlinear solver package (NOX) in the Trilinos software from SNL [3]. The nonlinear solve at each time step fully utilizes the JFNK framework in Trilinos with generalized minimum residual method (GMRES) as the linear solver. The linear solve required by the parabolic system for the implicit barotropic preconditioner is implemented with the conjugate gradient (CG) solver in Trilinos, which is in turn preconditioned with algebraic multigrid using the multilevel preconditioning package (ML) package in Trilinos. Results to date indicate that a stable, implicit time integration can be performed with step sizes thirty times larger than explicit methods with favorable scalability in terms of CPU time. Figures 1–3 show an example result from a prototype ocean problem.
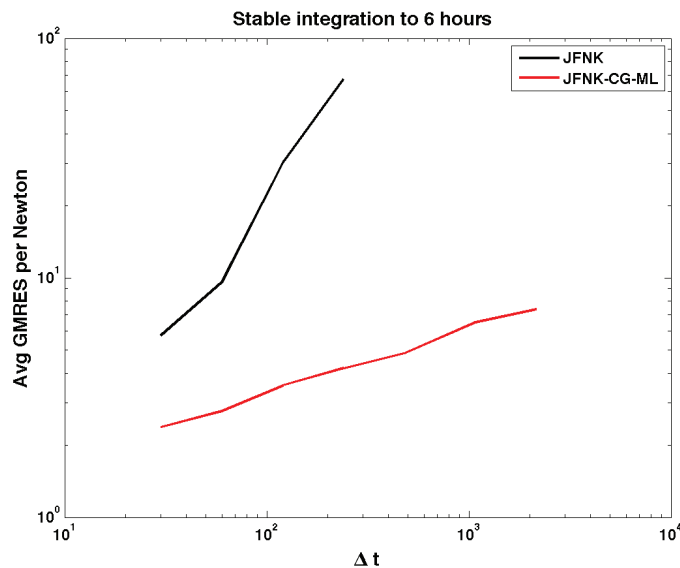
Fig. 2. Time-step size versus average number of GMRES iterations per Newton iteration for unpreconditioned (JFNK) and preconditioned (JFNK-CG-ML) implicit Euler. The dominant cost of JFNK is the number of linear GMRES iterations per Newton iteration. Preconditioning effectively reduces the number of linear GMRES iterations per Newton iteration.
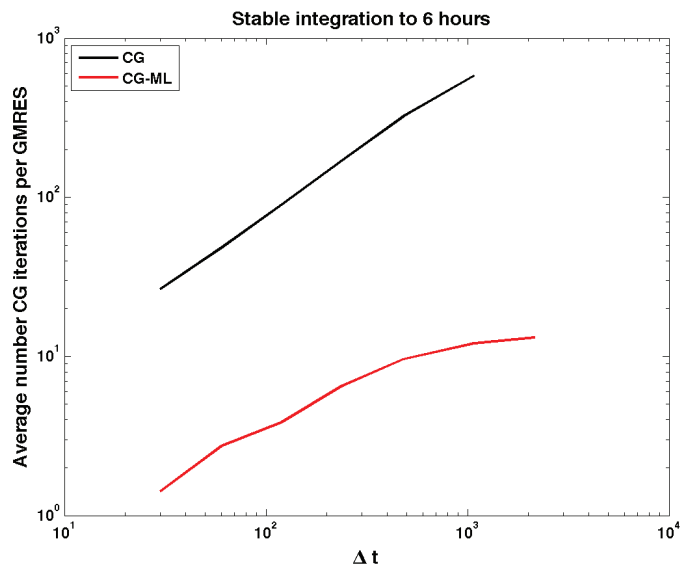
Fig. 3. Time-step size versus average number of CG iterations per GMRES iteration for inversion of the barotropic subsystem without preconditioning (CG) and with multigrid preconditioning (CG-ML). The dominant cost in application of the preconditioner is the number of linear iterations to invert the elliptic operator. Preconditioning of the CG iteration with multigrid methods results in a factor of 10 reduction in CG iterations and exhibits favorable scalability.

We are extending these ideas to co-design efforts in climate science at LANL. As baroclinic-barotropic decoupling is heterogeneous in nature, it readily fits a heterogeneous computing model. The fast-time-scale, 2D barotropic system can be implemented on traditional compute nodes, while the slow time-scale, 3D baroclinic system will be implemented on associated accelerated hardware. This model will allow for a paradigm shift in the use of heterogeneous computing, enabling the effective use of exascale resources. As can be seen, this algorithm shows strong potential in modern high-resolution climate simulations, allowing tighter coupling between the physics, thus reducing errors and increasing stability inherent to operator splitting. Having the ability to perform climate simulations with reduced splitting errors and using variable-resolution grids will be a significant step forward in predictive ocean and sea ice simulation. In addition, there is a clear route to exascale–given the heterogeneous nature of this algorithm, it can be readily mapped to evolving hybrid computing architectures, providing the opportunity for effective co-design.

[1] Knoll, D.A. and D.E. Keyes, *J Comput Phys* **193,** 357 (2004).
[2] Dukowicz, J.K. and R.D. Smith, *J Geophys Res* **99,** 7991 (1994).
[3] Heroux, M. et al., http://trilinos.sandia.gov (2008).

Improved Performance of MPAS-Ocean, an Unstructured-grid Ocean-Climate Model

Mark R. Petersen, CCS-2; Todd D. Ringler,
Douglas W. Jacobsen, Mathew E. Maltrud, Philip W. Jones, T-3

The Model for Prediction Across Scales-Ocean (MPAS-Ocean) is a new ocean model being developed for global climate change studies. MPAS-Ocean uses unstructured horizontal grids that enable it to perform high-resolution regional simulations within low-resolution global grids, with smoothly varying grid transitions. Major accomplishments this year include a split explicit time-stepping algorithm that improves performance by a factor of 12, automated performance testing, high-order horizontal and vertical advection, implicit vertical mixing, and an Arbitrary Lagrangian-Eulerian (ALE) vertical coordinate. MPAS-Ocean is now a fully functioning ocean model, and successful simulations have been run on realistic global quasi-uniform grids, variable resolution global grids down to 10-km gridcells, and a number of idealized test cases.

Fig. 1. Sample MPAS-Ocean unstructured grids. With MPAS, users have a great deal of freedom to specify mesh density, so computational resources can be spent where they are most needed.
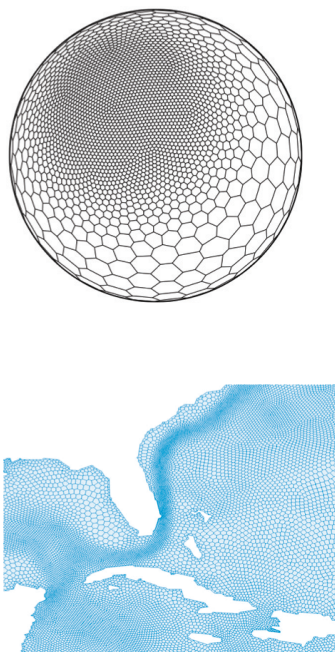
In order to prepare a new climate model component for public release and use in century-long climate-change simulations, there are several requirements that must be satisfied. The model must be well-documented and easy to run, standard parameterizations for vertical and horizontal mixing must be included, and the model must be validated against trusted metrics. In the end, however, a new climate-model component must be computationally efficient to be widely adopted by the community.

The Model of Prediction Across Scales-Ocean (MPAS-Ocean) is a new ocean-climate model developed by the Climate, Ocean, and Sea Ice Modeling (COSIM) team at LANL. It uses horizontally unstructured grids, so that high-resolution regions may be embedded within a lower-resolution global grid (Fig. 1). This is useful for climate-change studies of a particular region while capturing global processes that influence the area in question. The global meshes, created using Spherical Centroidal Voronoi Tesselations [1], vary smoothly from low- to high-resolution regions. The same algorithms and in fact, the same code, are used in all gridcells. This stands in contrast to traditional regional models, which use many nested uniform grids with sharp transitions in resolution, and even use separate models for these different regions. Important conservation properties for mass, tracers, potential vorticity, and energy are maintained, even when using variable resolution meshes [2,3]. The underlying scheme when paired with variable resolution meshes was also evaluated with encouraging results in 2011 [4].

MPAS is part of a collaborative effort between LANL and the National Center for Atmospheric Research (NCAR) in Boulder, Colorado, to create earth system model components for unstructured grids. The MPAS framework includes ocean, hydrostatic and nonhydrostatic atmosphere, and shallow water components and plans are in place for sea ice and land ice components.

The major effort with MPAS-Ocean in 2011 focused on improving model performance. Development of MPAS-Ocean began in 2010, and by year's end it was a functioning ocean dynamical core that could run globally with land boundaries and topography. However, the Runge-Kutta time-stepping algorithm in place at that time was too slow to be a viable method for climate change simulations. The solution was to advance to a split time-stepping scheme, which improved performance by a factor of 12.

The newly implemented time-stepping algorithm in MPAS-Ocean is a split explicit barotropic/baroclinic scheme. In ocean models, the time step is severely restricted by the fast surface gravity waves, which travel at ~200 m/s, while the next time-step limitation is imposed by slow internal gravity waves at 1–2 m/s. The surface wave dynamics may be solved efficiently using 2D barotropic (vertically integrated) equations at a small time step, while the remaining 3D baroclinic system, now free of surface waves, may take a large time step. In practice, the barotropic system is subcycled within each baroclinic time step. The whole process is repeated within a predictor-corrector scheme, and each stage includes individual iterations. In the end, there were many degrees of freedom to optimize in order to obtain a stable and efficient algorithm.

Many other additions were made to MPAS-Ocean to improve performance. The code was optimized by fusing loops and replacing branches with masks within loops, resulting in a 10% speed-up. The code was modularized to improve organization and readability, and new timers provide detailed performance profiles. Scripts were used to automate performance testing so that modifications could be
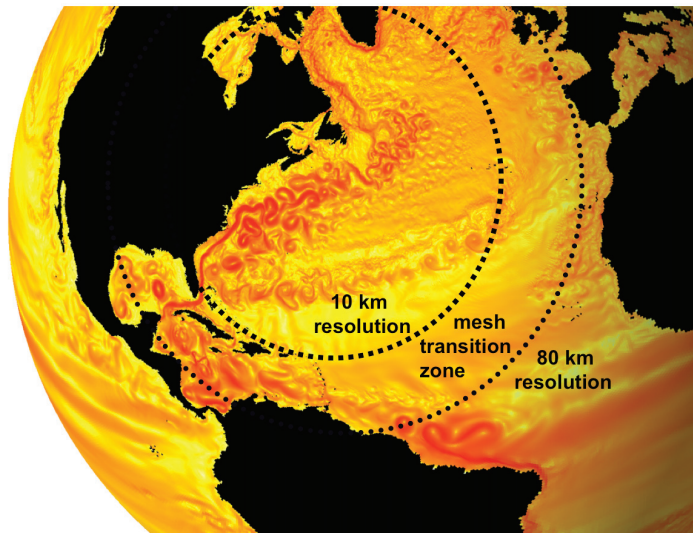
Fig. 2. Surface kinetic energy from a global, real-bathymetry simulation. The mesh density varies smoothly from 80 km over most of the globe, to 10 km in the North Atlantic. Significant eddy activity is visible in the high-resolution region.

incrementally tested for speed and accuracy over a variety of test configurations.

Performance tests using realistic global domains show excellent scaling out to thousands of processors and from resolutions of 120-km to 15-km gridcells. The progression from the Runge-Kutta time-stepping to the split explicit time-stepping shows a 12-fold speed-up. These tests were performed on Lobo, an Institutional Computing resource at LANL.

POP, the Parallel Ocean Program, is an ocean-climate model developed and maintained by the COSIM team since 1992. It has been used extensively for climate-change simulations that are published in the Intergovernmental Panel on Climate Change (IPCC), and so is a good comparison for MPAS-Ocean performance. MPAS-Ocean is currently 1.9 times slower than POP at lower resolution, and 3.4 times slower at higher resolution (compare the blue and red lines on Fig. 2). POP runs on structured rectangular grids, so it has the advantage that horizontal neighbors are brought into local memory in the same fetch. MPAS-Ocean's unstructured grid means that horizontal neighbors are accessed through pointers, and may not be nearby in memory. To compensate for this, MPAS-Ocean orders tracers and vertical neighbors first in memory, and does not include land gridcells.

Considering that POP has been scrutinized for performance improvement for the last 20 years and MPAS-Ocean uses completely different data structures and time splitting, we feel that a factor of 1.9-3.4 in speed between POP and MPAS-Ocean is a major accomplishment. We are confident that MPAS-Ocean performance can be improved even further. The number of messages passed among CPUs can be reduced by packaging many variables into a single message, and using very wide

halos in the barotropic variables that are explicitly subcycled. Tracers and density may be updated less frequently than the dynamic variables with little loss in accuracy.

There have been many accomplishments in MPAS-Ocean development in 2011 other than performance. High-order horizontal and vertical advection has been added for accurate tracer transport. Richardson-number-based vertical mixing and implicit vertical mixing are required so that vertical diffusion increases in less stratified conditions, and so the time step is not restricted when high vertical diffusion is required. We have implemented an Arbitrary Lagrangian-Eulerian (ALE), or generalized vertical coordinate, so that vertical gridcell interfaces oscillate with changes in sea-surface height, reducing spurious vertical mixing between layers.

Global simulations are now in progress using high-resolution quasi-uniform meshes and variable-resolution meshes. Figure 2 shows results from a North Atlantic regional variable-resolution mesh, where significant eddy activity can be seen in the Gulf Stream within the 10-km resolution region. The low-resolution 80-km region does not resolve eddies, but provides global currents that drive the North Atlantic region of interest.

In the coming year, we plan to introduce MPAS-Ocean to the ocean modeling community in several publications, and couple MPAS-Ocean to the Community Earth System Model (CESM) to run earth-climate simulations that include a dynamic atmosphere, sea ice, and land surface. A longer-term research goal is to develop "scale-aware" parameterizations for horizontal mixing that work well on variable resolution meshes without the need for tuning of coefficients on each new grid.

[1] Ringler, T. et al., *Ocean Dynam* **58**, 475 (2008).
[2] Thubern, J. et al., *J Comput Phys* **228**, 8321 (2009).
[3] Ringler, T. et al., *J Comput Phys* **229**, 3065 (2010).[4] Ringler, T. et al., *Mon Weather Rev* **139**, 3348 (2011).

The Atlantic Meridional Overturning Circulation and its Sensitivity to Enhanced Runoff from the Greenland Ice Sheet

Wilbert Weijer, CCS-2; Mathew E. Maltrud, T-3;
Matthew W. Hecht, CCS-2; Henk A. Dijkstra,
Michael A. Kliphuis, Institute for Marine and Atmospheric
Research Utrecht, The Netherlands

The Atlantic Meridional Overturning Circulation (AMOC) is an important component of the climate system, because it transports significant amounts of heat to high northern latitudes. The sensitivity of the AMOC to freshwater capping in the northern North Atlantic is a big concern in view of the current mass deficit of the Greenland Ice Sheet (GrIS), which is expected to accelerate in future decades. Determining the sensitivity of the AMOC to enhanced melt water from the GrIS is therefore essential for climate change projections.

In this study we address the sensitivity of the AMOC to enhanced run-off from the GrIS by comparing the AMOC response to enhanced melt-water input in a global, strongly-eddying ocean model to that in a non-eddying model typical of the current generation of climate models. We find that the AMOC response in the strongly-eddying model is greater than in the non-eddying model, suggesting that the AMOC may be more sensitive than previously thought.

The Atlantic Meridional Overturning Circulation (AMOC) is an important component of the climate system, as it transports significant amounts of heat to high northern latitudes [1]. Its deep branch consists of the southward transport of relatively cold North Atlantic Deep Water (NADW, roughly between 1000 and 2000 m), and is compensated by a northward flow of warmer waters at shallower levels. The production of NADW takes place mainly in the Labrador and Nordic Seas by a process called deep convection. Convection is induced by intense surface cooling (which tends to make the surface waters less buoyant), but is vulnerable to freshwater input (which tends to make the surface waters more buoyant). This is a serious concern considering the fact that the Greenland Ice Sheet (GrIS) is losing mass at an increasing rate due to the warming of the climate system [2]. Assessing the sensitivity of the AMOC to melt water from the GrIS is therefore an important task for climatologists.

Unfortunately, water mass transformations are notoriously hard to represent in the current generation of numerical climate models—spatial resolution is sacrificed for the requirement to integrate these models for centuries, and the ocean's turbulent eddy field is not explicitly represented but parameterized. In order to study how those turbulent processes influence the sensitivity of the AMOC to enhanced melt water input, we performed a suite of integrations with two configurations of the same ocean model [3]. The strongly eddying configuration (indicated by $R_{0.1}$) has a high spatial resolution that is fine enough to resolve important transport processes like eddies, narrow and energetic boundary currents, and small-scale convective processes. The non-eddying configuration (called x1) has a low spatial resolution that is characteristic of the current generation of climate models; the transports accomplished by eddies are necessarily represented by complex parameterizations.

In addition to simulations without any anomalous forcing we performed sensitivity experiments where we applied an anomalous freshwater flux around Greenland to mimic enhanced run-off and calving from its extensive ice sheet. We found that there is a significant difference in the response of the Meridional Overturning Circulation (MOC) strength (Fig. 1): in the non-eddying case there is a strong initial slow-down of the MOC, followed by a gradual adjustment to a new equilibrium. In the strongly eddying configuration, however, the MOC declines more gradually but for a much longer period of time, with the final decline being almost double that in the non-eddying model. Apparently the sensitivity of the MOC to a freshwater flux perturbation is enhanced when smaller-scale transport processes are explicitly resolved.

A main difference between the two model configurations is the representation of transport processes. In the strongly-eddying case, transports by turbulent eddies are explicitly accounted for, whereas they are parameterized in the non-eddying model. To monitor how this affects the dispersion of the freshwater signal, we put a dye tracer into
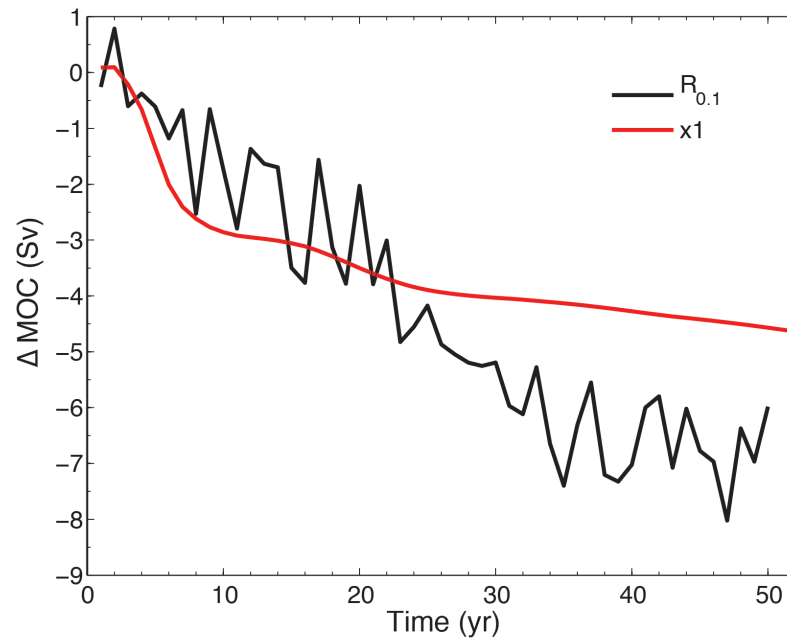
Fig. 1. Decline in the strength of the AMOC in response to enhanced melt water input from the GrIS, measured with respect to a control simulation (1 Sv = 10^6 m^3/s). Red line is for the non-eddying model configuration, which is characteristic of the current generation of climate models, black line is for the strongly eddying configuration.

the ocean around Greenland along with the melt water. Figure 2 shows the time it takes for this dye to reach a certain location in the Atlantic Ocean, either close to the surface (upper panels) or at depth (lower panels). It is clear that the dye is transported southward much more quickly when eddies are doing the job. However, there are also shadow zones (specifically in the eastern off-equatorial Atlantic) where no dye has penetrated even after the full 50 years of integration.

The main conclusion is that explicitly resolving turbulent transport processes matters for the sensitivity of the AMOC to melt water fluxes, and that the AMOC may be more sensitive than previously thought [4].
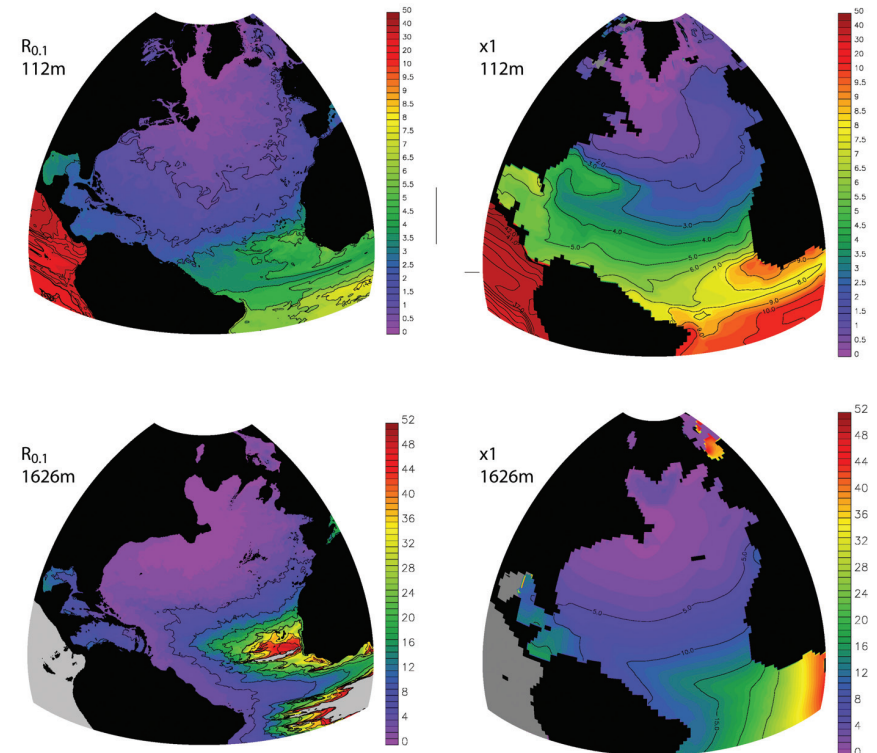
Fig. 2. Arrival time (in years) for a dye introduced along with the melt water around Greenland to reach a certain location in the Atlantic. Left column is for the strongly eddying configuration, right for the non-eddying case.

[1] Broecker, W.S., *Oceanography* **4,** 79 (1991).
[2] van den Broeke, M. et al., *Science* **326**, 984 (2009).
[3] Maltrud, M. et al., *Environ Fluid Mech* **10,** 275 (2010).
[4] Weijer, W. et al., *Geophys Res Lett,* submitted (2011).

Los Alamos National Laboratory Associate Directorate for Theory, Simulation, and Computation (ADTSC) LA-UR 12-20429
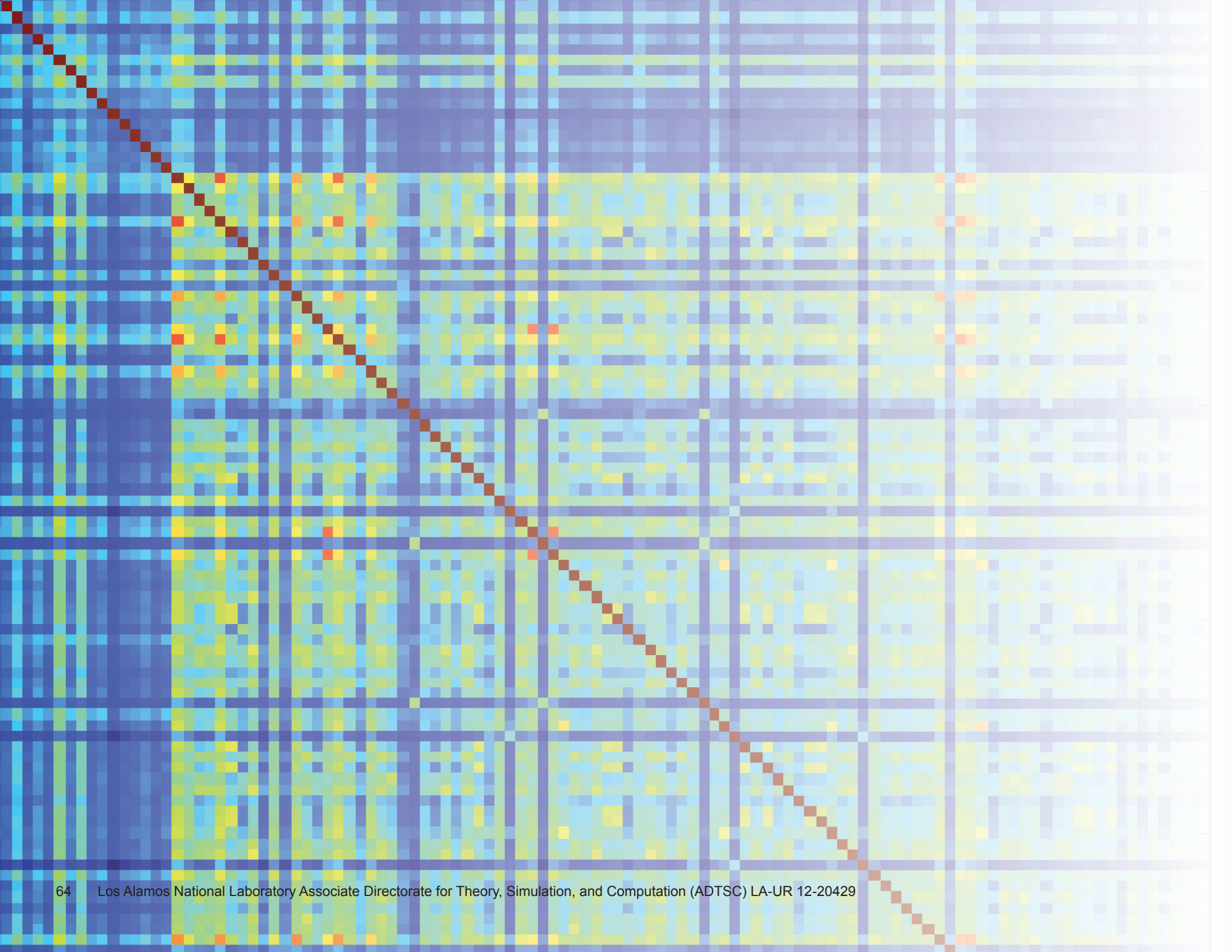
Information Science and Technology

The Information Science and Technology field draws upon computer science, computational science, and mathematics to provide innovation in the areas of data processing and management and derive actionable information from data. Information Science and Technology capabilities underlie most science endeavors at LANL.

This section includes contributions to the field of Information Science and Technology in the broad area of applied computer science. There are four papers on various approaches for improving I/O performance, a vital area of research as the gap between disk capacity and disk bandwidth grows. There are three papers on algorithms covering a broad spectrum from cybersecurity to optimizing selection on a GPU. There is also a paper that describes a quantitative approach to co-design for applications in molecular dynamics.

Graph-based Malware Detection Using Dynamic Analysis

**Blake H. Anderson, Daniel A. Quist, Joshua C. Neil,
ACS-PO; Curtis B. Storlie, CCS-6; Terran Lane,
University of New Mexico, Michael E. Fisk, ACS-PO**

We introduce a novel malware detection algorithm based on the analysis of graphs that are constructed from dynamically collected instruction traces of the target executable. These graphs represent Markov chains, where the vertices are the instructions and the transition probabilities are estimated by the data contained in the trace. We use a combination of graph kernels to create a similarity matrix between the instruction trace graphs. The resulting graph kernel measures similarity between graphs on both local and global levels. Finally, the similarity matrix is sent to a support vector machine to perform classification. Our method is particularly appealing because we do not base our classifications on the raw n-gram data, but rather use our data representation to perform classification in graph space. Our results show a statistically significant improvement over signature-based and other machine-learning-based detection methods.

Malware continues to be an ongoing threat to modern computing. In 2010, more than 286 million unique variants of malware were detected [1]. Despite the majority of this new malware being created through polymorphism and simple code obfuscation techniques, and thus being very similar to known malware, it is still not detected by signature-based anti-virus programs [2]. With the ever-increasing proliferation of these threats, it is important to develop new techniques to detect and contain these malware.

Many of the current anti-virus programs available rely on a signature-based approach to classify programs as being either malicious or benign. Signature-based approaches are popular due to their low false positive rate and low computational complexity on the end host, both of which are appealing for daily usage. Unfortunately, these schemes have been shown to be easily defeated by simple code obfuscation techniques [2]. With the ease of creating a new virus through these techniques and polymorphic viruses becoming more prevalent, non-signature-based methods are becoming more attractive.

To combat these issues, several researchers began to look at less strict measures to detect malicious code. These methods have generally revolved around n-gram analysis of the static binary or dynamic trace of the malicious program [3]. These methods have shown great promise in detecting zero-day malware, but there are drawbacks related to these approaches. The two parameters generally associated with n-gram models are n, the length of the subsequences being analyzed, and L, the

number of n-grams to analyze. For larger values of n and L, there is a much more expressive feature space that should be able to discriminate between malware and benign software more easily. But with these larger values of n and L, we run into the curse of dimensionality: the feature space becomes too large and we do not have enough data to sufficiently condition the model. With smaller values of n and L, the feature space becomes too small and discriminatory power is lost.

For our research, we use a modified version of the Ether Malware Analysis framework [4] to perform the data collection. Ether is a set of extensions on top of the Xen virtual machine. Malware frequently uses self-protection measures to thwart debugging and analysis. Ether uses a tactic of zero modification to be able to track and analyze a running system. Zero modifications preserve the sterility of the infected system, and reduce the methods that malware authors can use to detect if their malware is being analyzed. Increasing the complexity of detection makes for a much more robust analysis system. We use these modifications to allow for deeper introspection of the application programming interface (API) and import internals [5].

Our data representation gets away from the need to specify the appropriate n and L. Instead we model the dynamic instruction trace as a Markov chain represented by a weighted, directed graph. The instructions of the program are represented as vertices, and the weights of the edges are the transition probabilities of the Markov chain, which are estimated using the program trace we collect. An example

Fig. 1. A hypothetical fragment of a full Markov chain graph. In a real Markov chain graph, all of the outgoing edges would sum to 1.
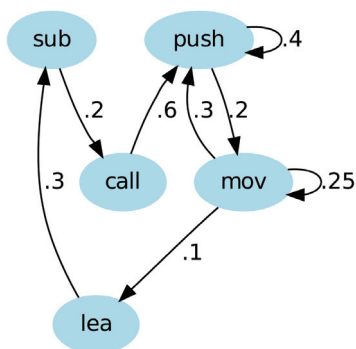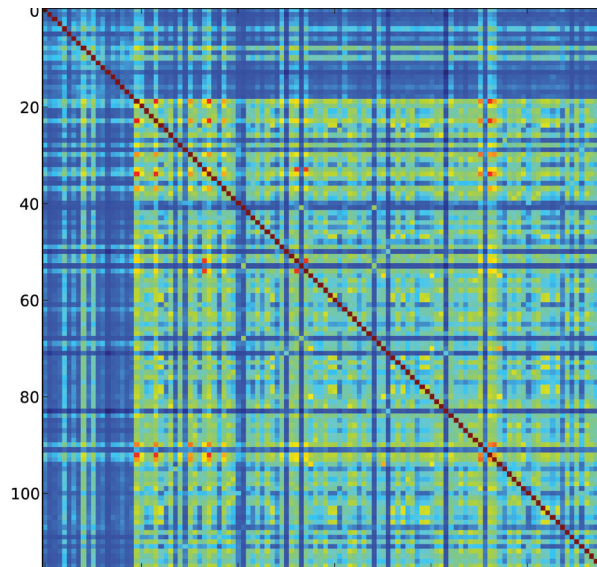
Fig. 2. The heat maps of the kernel (similarity) matrix for benign software versus malware. The smaller block in the upper left of the figure represents benign software and the larger lower right block represents malware.

of a hypothetical Markov chain graph is presented in Fig. 1. The novel contribution we present is to construct a similarity, or kernel, matrix between the Markov chain graphs and use this matrix to perform classification.

Our approach makes use of two types of kernels: a Gaussian kernel and a spectral kernel. The notions of similarity that these two kernels measure are quite distinct, and we found them to complement each other very well. The Gaussian kernel searches for local similarities between the adjacency matrices. It works by taking the exponential of the squared distances between corresponding edges in the weighted adjacency matrices. The motivation behind this kernel is that two different classes of programs should have different pathways of execution, which would result in a low similarity score.

The other kernel we use is based on spectral techniques [6]. These methods use the eigenvectors of the graph Laplacian to infer global properties about the graph. These eigenvectors encode global information about the graph's smoothness, diameter, number of components and stationary distribution, among other things. We then construct our second kernel by using a Gaussian kernel on these eigenvectors.

If we have two valid kernels, we are assured that a linear combination of these two kernels is also a valid kernel. This algebra on kernels allows us to elegantly combine kernels that measure very different aspects of the input data. We used cross-validation to find a suitable linear combination of our two kernels where the weights of the kernels are constrained to sum to one. Figure 2 shows a heat map of the combined kernel matrix with 19 instances of benign software and 97 instances of

malicious software. Once the combined kernel matrix is constructed, we use a standard support vector machine to perform classification.

To validate our approach, we used a dataset composed of 615 instances of benign software and 1,615 instances of malware. We tested our combined kernel, Gaussian kernel, spectral kernel, and the n-gram methodology using a support vector machine, values of n ranging from two to six, and values of L ranging from 500 to 3,000 in increments of 500. We also tested against 10 different signature-based anti-virus programs. Our combined kernel was the overall winner with an accuracy of 96.41%. The best n-gram method with n=3, L=2500 had an accuracy of 82.15%. It is interesting to note that all machine-learning-based methods easily beat the best signature-based anti-virus, which had an accuracy of 73.32%.

Our novel method extends the n-gram methodology by using 2-grams to condition the transition probabilities of a Markov chain, and then treats that Markov chain as a graph. Taking the Markov chain as a graph allows us to utilize the machinery of graph kernels to construct a similarity matrix between instances in our training set. We use two distinct measures of similarity to construct our kernel matrix: a Gaussian kernel, which measures local similarity between the graphs' edges, and a spectral kernel, which measures global similarity between the graphs. Given our kernel matrix, we can then train a support vector machine to perform classification on new test data.

[1] Symantec White Paper, *Internet Security Threat Report* **16,** (2011).
[2] Christodorescu, M. and Jha, S., *Proc 12th USENIX Secur Symp* (2003).
[3] Reddy, D. and A. Pujari, *J Comput Virol* **2,** 231 (2006).
[4] Dinaburg, A. et al., *Proc 15th ACM Conf Comput Comm Security* (2008).
[5] Quist, D. et al., *J Comput Virol* **7,** 121 (2011).
[6] Chung, F.R.K., *Spectral Graph Theory,* American Mathematical Society (1997).

A New Data Smoother Applied to Low-resolution Sodium-Iodide Detectors

Thomas L. Burr, CCS-6;
Nicholas W. Hengartner, CCS-3;
Steven C. Myers, N-2

The energy spectra of gamma-rays emitted by radioisotopes act as fingerprints that enable identification of the source. Such identification from low-resolution sodium iodide (NaI) detectors over short time periods is challenging for several reasons, including the Poisson fluctuations in the recorded counts. Smoothing the data over neighboring energy bins can reduce noise in the raw counts, at the cost of introducing a bias that de-emphasizes the peaks and valleys of the spectrum. This note describes a new two-stage smoothing procedure that uses a multiplicative bias correction for adjusting initial smoothed spectra. The benefits of this new method are illustrated on real spectra.

Performance of radioisotope identification (RIID) algorithms using gamma-ray spectroscopy is a subject of increasing importance. For example, sensors deployed at locations that screen for illicit nuclear material rely on isotope identification to resolve innocent nuisance alarms arising from naturally occurring radioactive material (NORM) [1-3]. Studies comparing the energy spectra measured by low, medium, and high resolution detectors suggest that relatively low-resolution detectors, such as sodium iodide (NaI) detectors that typically count the number of gamma photons in 512 or 1024 calibrated energy channels, will continue to play a key role in NORM identification [1]. However, RIID performance from the raw counts collected over short time intervals is strongly impacted by intrinsic variations from Poisson count statistics.

Various strategies for improving the RIID by reducing the noise in the data are being considered. One noise-reduction strategy is to smooth the spectral data over neighboring energy bins. This smoothing decreases the variance of the raw counts at the cost of introducing some bias that de-emphasizes the peaks and valleys of the spectrum. Most smoothers also distort peak shape by broadening the base and midsection. Preliminary results in [4] suggest, but do not prove, that substituting the raw NaI spectral counts with their smoothed values can lead to improved performance of existing RIID algorithms. Even better performance can be expected if one can reduce the bias of the smoothed counts without substantially increasing their variance.

Reference [5] describes a new two-stage smoother that uses a multiplicative bias correction (MBC) for adjusting initial smoothed energy spectra, thereby correcting some of the deficiencies of traditional smoothers. Reference [6] describes the effect of the MBC-smoother in the context of detecting injected point-like threat isotopes onto each of many real vehicle profiles.

The bias of main concern in smoothing spectral data is the tendency by all smoothers (local polynomial, local kernel, wavelet, etc.) to underestimate the peaks and overestimate the valleys of the true spectrum f(x). The true count rate as a function of energy x is denoted f(x), and the observed count rate in a given spectrum is denoted $y = f(x) + e$ (see Fig. 1). High-degree-of-freedom fits, such as the 100-degree-of-freedom spline in Fig. 1, will have smaller bias in the peaks, but larger variance in the off peaks. As a result, the peaks are fit very well, but the smoother tends to generate spurious smaller peaks. The opposite is true for low-degree-of-freedom fits, in which the smoother fails to fully catch the peaks but does not tend to introduce additional spurious peaks.

Finding a smoother that minimizes the root mean squared error (RMSE) does not by itself guarantee that the resulting smoother has desirable properties at the peaks while not generating spurious ones. This leaves open the question of selecting the smoothing parameter. Instead of attempting to optimize the bandwidth h for the RIID application, we focus on estimating the bias and correcting the initial smoother using that estimate. This is not a new idea, but the application is new, and most of the approaches described in the literature focus on additive corrections of the bias. Such corrections may not be desirable in the present context, as they can lead to negative-valued energy spectra that are undesirable as input into existing RIID algorithms.

Fig. 1. Example ^{232}Th spectrum for one minute, for sixty minutes (top), and two example spline fits (bottom). The 100 df spline over-fits the off-peak channels. The 35 df spline over-smooths the peaks and valleys.

INFORMATION SCIENCE AND TECHNOLOGY

For more information contact Thomas L. Burr at tburr@lanl.gov.
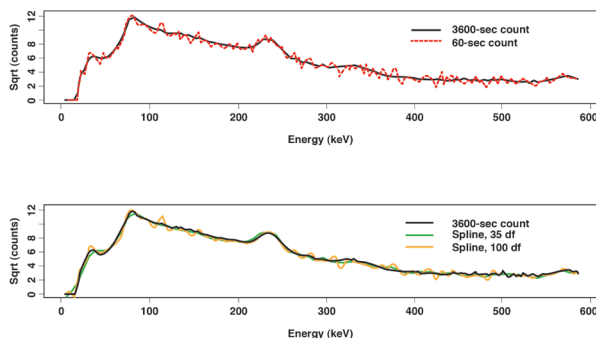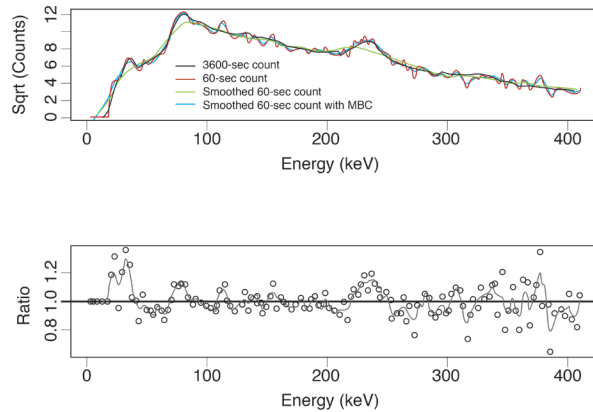
Fig. 2. Example ^{232}Th spectrum for 3600 seconds, for 60 seconds, and a lokerns fit with and without the multiplicative bias correction; (bottom) The ratio $R_i = \dfrac{y_i}{\hat{y}_{1,i}}$

A recent attractive alternative [5 is to apply an MBC to an initial smoother. The asymptotic analysis and small sample simulations in [7] show that this method is effective at reducing the bias of the initial smoother without increasing the variance. As a result, one obtains estimates for the smoothed spectra that better fit the peaks and valleys but do not introduce additional spurious peaks or under-smooth the flat regions. For good performance, the MBC approach needs the initial smoother to not under-smooth. Over-smoothing by the initial smoother is acceptable. Therefore, we used cross-validation to select the bandwidth for the initial smoother. Cross-validation [8] involves leaving out x,y pairs while calculating the smooth function at trial bandwidths and then testing the performance of the trial bandwidths on the held-out pairs.

For completeness, we describe the MBC smoother here. Starting with an initial kernel smoother for $\sqrt{f(x)}$, with weight function $w_{1,i}(x,h_1)$ and bandwidth h, $\hat{g}(x) = \sum_{i=1}^{n} w_{1,i}(x,h)\sqrt{y_i}$ we denote by $\hat{y}_{1,i} = \hat{g}(x_i)^2$ the predicted spectrum in bin i, and construct the auxiliary variables $R_i = \dfrac{y_i}{\hat{y}_{1,i}}$.

Each value Ri is an unbiased, but noisy, estimate of the relative bias in the ith energy bin. Smoothing the latter gives rise to a multiplicative correction factor $\hat{R}(x) = \sum_{i=1}^{n} w_{2,i}(x,h_2)R_i$ for the initial smoother.

Multiplying the predicted spectrum by the latter leads to the estimate $\hat{f}_2(x) = \hat{R}(x)\hat{g}(x)^2$ for the energy spectra f(x), and in particular the estimate $\hat{y}_{2,i} = \hat{f}_2(x_i)$ for the expected counts of gamma-photons in the ith bin. Neither the bandwidth nor the smoother for the correction factor need be the same as for the initial smoother.

Figure 2 illustrates the MBC for the same ^{232}Th source as in Fig. 1. Notice again at least three distinct peaks and notice the trend in $R_i = \dfrac{y_i}{\hat{y}_{1,i}}$, which suggests that peaks will be better estimated using the corrected values $\hat{y}_{2,i}$ rather than predictions from the initial smoother $\hat{y}_{1,i}$.

In NaI spectra, peak width varies as a function of energy because detector resolution varies with energy. Adaptive smoothing is therefore preferred. There are two main options for the adaptive smoothing of y = f(x) + e. One option is to adjust the local bandwidth according to the local density of the x data. This option is not relevant for our context because of equally spaced x values, regardless of whether one regards x to be energy or channel number. A second option is to identify regions of rapid change in y, such as peak regions, and modify the bandwidth accordingly. In effect, the MBC applies the second option. However, note that the MBC does not require one to identify the peak regions. This is a key advantage of the MBC, which provides adaptive smoothing, by "smoothing less" in the peaks. Smoothing less in the peaks occurs because the ratio $R_i = \dfrac{y_i}{\hat{y}_{1,i}}$ used in the MBC tends to pull the first smooth $\hat{y}_{1,i}$ toward y_i and hence leads to "smoothing less," as desired in the peaks.

We have found that both smoothing splines and a local kernel smoother (such as lokern) work well as a second smoother, although that choice is not critical.

[1] Stromswold, D. et al., IEEE Nucl Sci Symp, http://ieeexplore.ieee.org/iel5/9892/31433/01462180.pdf (2004).
[2] Burr, T. and M.S. Hamada, *Algorithms* **2,** 339 (2009).
[3] Burr, T. et al., *Qual Reliab Eng Int* **25**, 899 (2009).
[4] Myers, S. et al., "Choosing the Number of Repeat Measurements in Radioisotope Identification Performance Testing," Los Alamos National Laboratory Unrestricted Release Report LA-UR 08-1764 (2008).
[5] Burr, T. et al., *IEEE Trans Nucl Sci* **57**, 2831 (2010).
[6] Burr, T. et al., *Appl Radiat Isot* **69,** 1436 (2011).
[7] Hengartner, N. et al., "Multiplicative Bias Corrected Nonparametric Smoothers with Application to Nuclear Energy Spectrum Estimation," http://arxiv.org/abs/0908.0128v1 (2009).
[8] Hastie, T. et al., *The Elements of Statistical Learning, Springer,* New York (2001).

A Prototype of Exascale Checkpoint and Restart Computing Environment Using PLFS and Burst-Buffer Supports

Hsing-bung Chen, HPC-5; Gary A. Grider, HPC-DO;
Parks M. Fields, Josip Loncaric, HPC-5;
David R. Montoya, HPC-3

LANL has designed and developed a scalable and high performance Parallel Log-Structure File system (PLFS) for checkpoint and restart operations. PLFS has demonstrated its superior handling capability with parallel "N-to-1" file input/output (I/O) problems. Here we present a prototype Exascale checkpoint and restart computing environment using LANL's PLFS and local solid state device (SSD)-based burst-buffer storage enhancement supports. We use the global name space capability of the parallel file system as a metadata backend and the high-speed-data bandwidth of the local burst-buffer storage pool as a data backend. We illustrate a scalable checkpoint and restart system to meet Exascale computing requirements. We also propose three new collective MPI-IO application programming interfaces (API), MPI_File_protect_all, MPI_File_ensure_all, and MPI_File_delete_all, to support this prototype implementation, and use these to provide a space management sub-system between the burst-buffer and parallel file systems.

According to the technology roadmap of future Exascale computing systems, we expect checkpoint and restart file input/output (I/O) bandwidth to be in the range of 20 TB/sec to 60 TB/sec. Traditional disk-based parallel file systems cannot meet these challenging requirements in terms of cost, power consumption, reliability, and data access bandwidth. To meet these requirements for exascale computing, we need to introduce an additional layer to the traditional computer memory hierarchy. New storage-class memory technologies, such as solid state devices (SSD) and phase-change memory (PCM), provide very promising high-data bandwidth, high I/O operations per second (IOP) and low data-access latency.

We intend to leverage these new storage technologies and use them as a localized read/write burst-buffer between memory and the traditional disk-based parallel file system. The combination of using a storage-class memory-based local burst buffer and disk-based parallel file system gives us a new prospect for the exascale checkpoint and restart computing environment.

We first write checkpoint data to the local burst-buffer and later asynchronously migrate important long-term data from the burst buffer to a parallel file system. The burst buffer gives us memory-scale storage performance and, at the same time, disk-based parallel file systems can sustain the large storage demands from applications.

LANL has designed and developed a scalable and high-performance Parallel Log-Structure File system (PLFS) for checkpoint and restart file I/O operations. PLFS has already demonstrated its superior handling capability with parallel "N-to-1" file I/O problems. In this project, we present a prototype of the exascale checkpoint and restart computing environment using LANL's PLFS and local SSD-based burst-buffer storage enhancement supports. We utilize the global name space capability of the parallel file system as a metadata backend and the high-speed data bandwidth of the local burst-buffer storage pool as a data backend. We illustrate a scalable checkpoint and restart system to meet the exascale computing requirement. We also propose three new collective MPI-IO application programming interfaces (API): MPI_File_protect_all, MPI_File_ensure_all, and MPI_File_delete_all, to support this prototype implementation. We use these three new APIs to provide a space management sub-system between the burst-buffer and parallel file systems.

Fig. 1. A hybrid storage system: bridging capacity and bandwidth for exascale computing systems.

Fig. 2. A prototype of exascale checkpoint and restart system using hybrid storage system.

Fig. 3. Proposed three MPI-IO APIs to support burst-buffer checkpoint and restart operations.

Indago: A Novel Search and Analysis Tool for Electronically Stored or Transmitted Content

Jorge H. Román, HPC-1; David H. DuBois, HPC-5;
Shelly Spearing, HPC-1; Andrew J. DuBois,
Mike Boorman, HPC-5; Ekaterina A. Davydenko,
HPC-1; Carolyn M. Connor, HPC-5; Robert L. Gurule,
HPC-1

As data generation and storage technologies have advanced, society itself has become increasingly reliant upon electronically generated and stored data. Digital content is proliferating faster than humans can consume it. Current search/filter technology is well suited for simple searches/matches, but a more powerful paradigm is required for complex searches such as finding sensitive corporate knowledge that may be flowing in the intranet and could be accidentally or maliciously sent out over the internet. Context-based search and analysis can greatly enhance the exploration of data and phenomena by essentially reducing the data deluge and increasing the efficiency and effectiveness of the human analyst and/or end-user to access and fully exploit the data-to-knowledge potential that is inherent but latent in nearly every collection.

As technology has continued to advance, modern society has become increasingly reliant upon electronically generated and stored data and information. Digital archives are growing everywhere both in number and in size. Correspondingly, the need to process, analyze, sort, and manipulate data has also grown tremendously. Researchers have estimated that by the year 2000, digital media accounted for just 25% of all information in the world. After that, the prevalence of digital media began to skyrocket, and in 2002, digital data storage surpassed non-digital for the first time. By 2007, 94% of all information on the planet was in digital form [1]. The task of processing data can be complex, expensive, and time-consuming. Applications that alleviate the processing burden and allow users to access and manipulate data faster and more effectively to cross the data-to-knowledge threshold to enable informed, actionable decision-making, particularly for large data streams or digital repositories, are in demand.

Fig. 1. Graphical illustration of complex, context-based, relationships and connectivity of decision rule set.

Indago's primary function is the contextual analysis of large repositories of electronically stored or transmitted textual content. A key difference from current approaches is the weighted identification of "concepts in context." Most filters and search engines use either a simple word list or Boolean logic to express desired patterns. For example, Google.com uses a simple list, augmented by additional data (e.g., prestige of pages linking to the document, page ranking, etc.), to produce good retrieval rates; however, the number of matches can be impractically high with many false positives.

Indago's contextual analysis allows creation of complex models of the language contained in unstructured text in order to build increasingly sophisticated tools that move beyond word occurrence, enabling Indago to make connections and discover the patterns found in large collections.

Indago computes the relevance (or goodness-of-fit) of the text contained in an electronic document based on a predefined set of rules created by a subject matter expert. These rules are modular and are expressed as hierarchical concepts in context. The rules can be nested and are complex in nature to encode the subject matter expert's interpretation of a target concept, such as sensitive corporate knowledge. The system uses a combination of software and hardware to achieve near-real-time analysis of large volumes of text. The currently deployed implementation is used as a context filter for an email server. However, the technology has broad applicability, as the need for fast and accurate search, analysis, and/or monitoring of digital information transcends industry boundaries.

A second key difference from current approaches is that Indago provides a unique, hardware-assisted solution that exploits commercial off-the-shelf (COTS) hardware end-to-end. A fully operational software-based contextual analysis tool has been successfully deployed in a production

Screenshot of Indago output. Example analysis of a publically available news article (Dan Levin, "China's Own Oil Disaster," The Daily Beast, www.thedailybeast.com, July 27, 2010).

environment. In addition, a preproduction system has been developed to demonstrate hardware-accelerated capability.

The hardware-accelerated design enables the inspection of every bit of data contained within a document at rates up to 20 Gbps, far exceeding the capability of current generation multi-core processors while consuming substantially less energy, making Indago a high-performance, low power, green solution. Indago is an inline service as compared to implementations that require large computer clusters with high-end machines. Indago's current hardware acceleration is based on Netlogic technology. Netlogic technical specifications quote power consumption at ~1 Watt per 1 Gbps of processing speed; therefore, at the full rate of 20 Gbps of contextual processing, estimated power consumption would be 20 Watts, which is better than the power consumption of a single computer node by at least an order of magnitude. Comparison to a cluster of compter nodes, as some competing approaches require, is far more impressive. While this technology is designed for and widely applied in the data communications industry, Indago is the first to apply Netlogic's deep-packet processing technology to the field of digital knowledge discovery.

Indago is a cost-effective solution that provides unparalleled performance and capability using proven COTS. It allows users (e.g., scientists, information technology personnel, law firms, etc.) to engage their data faster, more accurately, and more effectively, thus allowing them to solve problems faster, more creatively, and more productively. Furthermore, Indago is domain-adaptable, power-efficient, and fully scalable in terms of rule set size and complexity.

[1] Martin, H. et al., *Science* **332,** 60 (2011).

iTransformer: Using SSD to Improve Disk Scheduling for High-Performance I/O

Kei Davis, CCS-7; Song Jiang, Xuechen Zhang,
Wayne State University

Data-intensive scientific computing applications are producing increasingly high Input/Output (I/O) demands on the storage devices of high-performance computing systems. Request concurrency, or the number of processes concurrently issuing requests, can be very high at data servers that are serving requests from applications running on a large-scale cluster. Besides the potentially large volume of requested data, this concurrency can significantly compromise the efficiency of a hard-disk-based storage system: data on a disk that are requested by different processes or programs are usually spatially separated on the disk, and concurrently accessing them can cause the disk heads to frequently seek from track to track, potentially delivering an I/O throughput an order of magnitude lower (or worse) than that for sequential disk access.

The emerging solid state drive (SSD) is largely unaffected by random access because it is, effectively, a uniform memory access device. However, it is currently not economical in high-performance computing for use as the main storage in a large-scale installation. More

cost-effective and practical options are either to use an SSD as buffer cache between main (dynamic random-access memory, DRAM) memory and the hard disk and exploit workloads' locality for data caching, or use SSD with a hard disk to form a hybrid storage device such that frequently accessed data is stored on the SSD.

These schemes for SSD usage, however, do not effectively address the problem of concurrent requests from data-intensive parallel programs. First, the data accessed in a single run of such a program can be larger than the capacity of the SSD. When a program processes a large data set, data are rarely accessed multiple times from the disk, and the accesses therefore exhibit weak temporal locality, which is hard to exploit for effective caching by a relatively small SSD. Second, requests to a disk are usually interleaved from different processes of one or multiple programs. Most existing SSD-based schemes exploit spatial locality, that is, they attempt placement of randomly accessed data on the SSD such that the hard disk serves requests of sequential, or at least well-ordered, data. However, when the request concurrency is high, it is highly likely that most requests from different processes will be presented to the disk as random access and will need to be handled by the SSD. This would overwhelm the SSD as a cache, or as a small storage device for random data, and make these schemes ineffective.

In the operating system, the I/O scheduler is the last opportunity to exploit spatial locality in the presence of high request concurrency. For example, CFQ, the default Linux disk scheduler, reduces random data accesses by merging and sorting outstanding requests according to their logical block addresses (LBA). Outstanding requests are kept in a data structure called a dispatch queue. The larger the queue, the more requests can be collected for sorting and the greater the chance to

Fig. 1. I/O throughput of a data server running IOzone with (a) 128 threads; and, (b) 256 threads. In each figure I/O throughputs with differing dispatch queue sizes and differing data access patterns are shown.

(a) 128 Threads

(b) 256 Threads
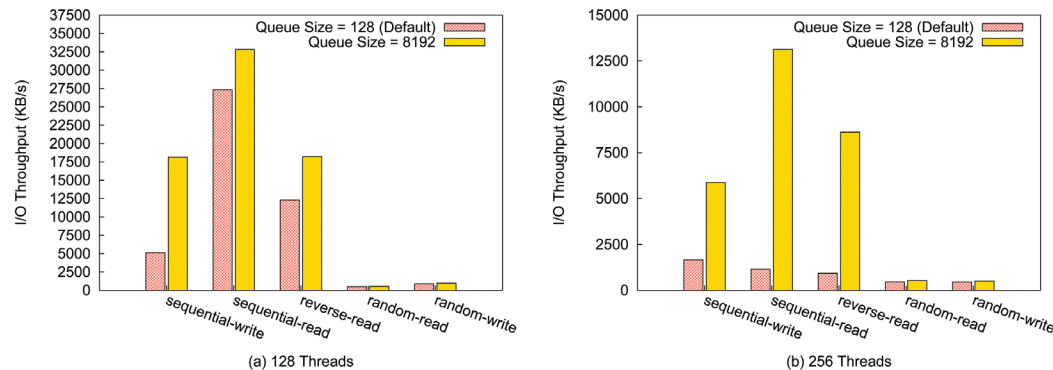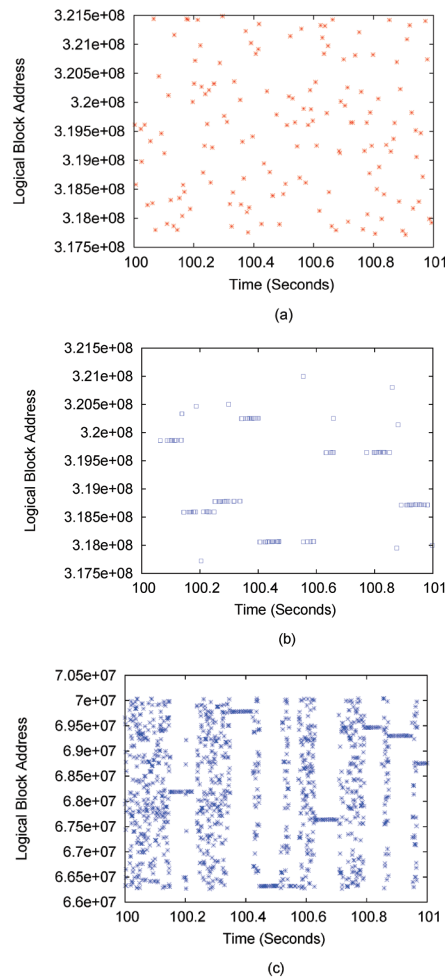
Fig. 2. Accessed locations when running ior-mpi-io using read requests in a sampled one-second execution period. (a) The locations are on the hard disk and the stock system is used. (b) The locations are on the hard disk and the system with iTransformer. (c) The locations are on the SSD and the system with iTransformer.

exploit spatial locality. The default queue depth in Linux's CFQ is 128.

We started by investigating the effect of simply increasing queue size on I/O performance in the presence of request concurrency by running IOzone, a commonly used file system benchmark, to generate a variety of file operations with varying queue size on a data server running Linux with CFQ to access data on a hard disk. Figure 1 shows I/O throughputs reported by the benchmark for access patterns Sequential Read/Write, Reverse Read, and Random Read/Write, with queue sizes 128 and 8192, with either 128 threads (Fig. 1a) or 256 threads (Fig. 1b).

This investigation showed that increasing queue size can significantly improve performance for Sequential Read/Write and Reverse Read. When the queue size is increased to 8192 the throughputs are significantly increased by 42% to 650%. This demonstrates that a large queue can effectively recover spatial locality if it exists in requests from the same thread. However, when individual threads issue fully random requests, the I/O throughputs are very low and the improvements made by the increased queue size are also small. This shows that random requests are at best difficult to schedule for efficient service by hard disk.

While increasing the size of the dispatch queue in memory can improve access locality for higher disk efficiency, by itself the approach has limitations. First, having a large queue would allow many write requests to be outstanding in volatile DRAM memory, running the risk of losing a large amount of data where frequent system failures are expected to be the norm. Second, although a long queue usually improves throughput, it can allow requests to remain in the queue for an extended period of time without being completed, which may result in excessive response times for those requests—for applications with strict quality of service (QoS) requirements a long queue can be

problematic. Third, as we showed in the experiments, simply increasing the queue size may not be sufficient, especially for addressing the issue of concurrency among streams of random requests.

Our design started with extending the scheduler's dispatch queue using SSD to hold the extension. Various algorithmic techniques were devised to overcome all of the aforementioned limitations. Our implementation of this scheduling architecture and the scheduling algorithm, collectively called iTransformer, is as a stand-alone Linux kernel module. The implementation is transparent to the software above the generic block layer in the kernel memory hierarchy and is therefore portable across different parallel file systems. Our evaluation of iTransformer was performed on the CSS-7 Darwin Cluster with a suite of representative benchmarks. We used 48 processes per node (one process per core on the DL585 nodes), one SSD per node, and the PVFS2 parallel virtual file system. iTransformer significantly reduced random access of the hard disks and increased I/O throughput of the storage system by up to three times and 35% on average, as compared to the stock system, for the benchmarks, using a mere 8 GB of each SSD. Figure 2 gives an example of how accesses are serialized to hard disk.

[1] Zhang, X. et al., "iTransformer: Using SSD to Improve Disk Scheduling for High-performance I/O," *26th IEEE Int Parallel Distr Process Symp,* to appear (2012).

Opportunistic Data-driven Execution of Parallel Programs for Efficient I/O Services

**Kei Davis, CCS-7; Song Jiang, Xuechen Zhang,
Wayne State University**

The work described here contributed to research into common runtime elements for programming models for increasingly parallel scientific applications and computing platforms. A parallel computing system relies on both process scheduling and input/output (I/O) scheduling to efficiently use resources and a program's performance hinges on the resource on which it is bottlenecked. Existing process schedulers and I/O schedulers are independent, but when the bottleneck is I/O there is an opportunity to alleviate it via cooperation between the I/O and process schedulers: the service efficiency of I/O requests can be highly dependent on their issuance order, which in turn is heavily influenced by process scheduling. We conceived a data-driven program execution model in which process scheduling and request issuance are coordinated to enable high I/O efficiency. Our design, DualPar, was implemented on the CCS-7 experimental cluster Darwin, and our experiments showed that DualPar can significantly increase system I/O throughput for relevant benchmarks [1].

Scientific computing is becoming increasingly data-intensive, with the consequence that the input/output (I/O) system is an increasingly severe performance bottleneck. In general, when a system resource becomes a parallel program's performance bottleneck, a better scheduling policy is sought to alleviate it. If the resource is the processors, this can be a process scheduling strategy for load balancing or co-scheduling. If the resource is storage, this could be an optimized I/O scheduler. In today's systems, when I/O service on the storage system becomes a program's bottleneck, the process scheduler becomes less relevant. This is especially true when I/O requests are mostly synchronous, most of the time the processes are idle waiting for the completion of their I/O requests and their scheduling is essentially a passive reaction to the progress of I/O operations.

Over decades the I/O stack, through which I/O requests pass and are serviced, has been significantly optimized, such as by forming larger sequential requests, hiding I/O time behind compute time with conservative prefetching,

Fig. 1. The architecture of DualPar.

or increasing the parallelism of data access with parallel file systems. However, in these scenarios the way in which processes are scheduled for execution is not considered for its effect on I/O efficiency. I/O requests are issued by processes and the requested data are consumed by processes. Therefore, the order in which the requests are issued and served can be significantly influenced by process scheduling. When a process is driven by its computation, the computation determines the request issuance order, which can directly affect the request service order and I/O efficiency. The throughput of a hard disk for serving sequential requests can be more than an order of magnitude higher than that for serving random requests, so I/O request issuance order is a critical factor in I/O efficiency.

We proposed and implemented a new data-driven execution mode for parallel programs that is enabled when I/O becomes a bottleneck and I/O efficiency is being compromised by request issuance order. In this mode, a process is scheduled to resume its execution not when the request on which it is currently blocked is completed, but when the data that it and its peer processes are anticipated to read has been prefetched into the buffer cache, or the data to write have been buffered in the cache. This then allows the processes to run longer before they block on a new I/O request. In the data-driven mode we not only require that requests of a process be served in a batch, but also that the serving of requests from different processes be coordinated. This is because requests from different processes may disruptively compete for disk service and degrade disk efficiency. Furthermore, this coordination
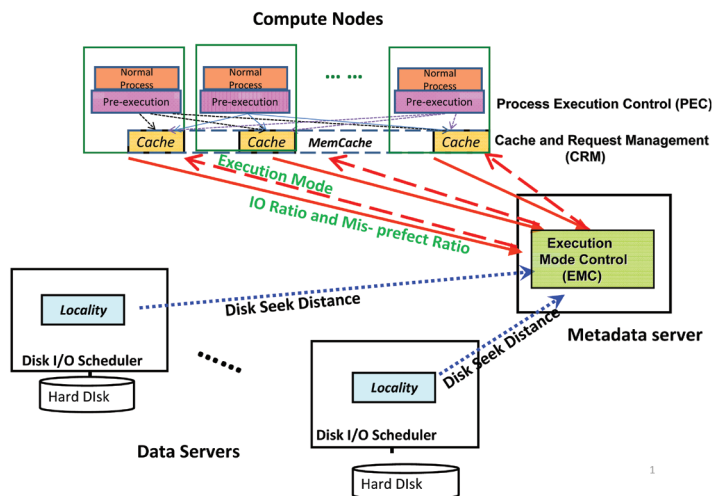
creates an opportunity to further improve the request issuance order to increase access sequentiality and to reduce the number of requests. Figure 1 illustrates the DualPar architecture.

In brief, we made the following contributions.

- We proposed a new program execution mode in which the scheduling of processes can be explicitly adapted for I/O efficiency. To this end, we use pre-execution to predict data to be requested for prefetching and a client-side buffer to hold written data for efficient writeback. In this way the computation of the program can be decoupled from the issuance of requests for its needed data, such that the I/O bottleneck can be alleviated, which cannot be achieved by conventional disk schedulers.

- We designed algorithms, which comprise DualPar, to detect the condition for enabling and disabling the data-driven mode and to coordinate data access and process executions.

- We implemented these algorithms in the MPICH2 MPI-IO library for message-passing interface (MPI) programs. We evaluated it with representative benchmarks, including mpi-io-test, ior-mpi-io, BTIO, and S3aSim. Experimental measurements on the CCS-7 large-scale experimental cluster Darwin show that DualPar can significantly improve I/O performance and, equally importantly, does not significantly degrade performance in other scenarios.

Figure 2 gives an example of relative I/O throughputs of standard MPI-IO, collective I/O, and DualPar for three concurrent instances of the BTIO benchmark. As shown, the advantage conferred by DualPar increases with the degree of parallelism.

Our basic premise was that, in the context of data-intensive computing, the independence of process scheduling and I/O scheduling, with their differing priorities, exacerbates the I/O bottleneck when using hard-disk-based storage. Preliminary experiments supported this premise, suggesting an opportunity to mitigate the I/O bottleneck via cooperation between the process and I/O schedulers. A new data-driven execution mode was conceived wherein processes would be scheduled according to the immediate (in-memory) availability of their I/O requests, this

in turn being enabled by the pre-execution of processes. This scheme, named DualPar, was implemented as part of the MPICH2 MPI-IO library. Extensive testing showed that DualPar could significantly improve the performance of I/O-bound parallel benchmarks, whether or not they used collective I/O. DualPar is transparent to the user, requiring no changes to application source code.
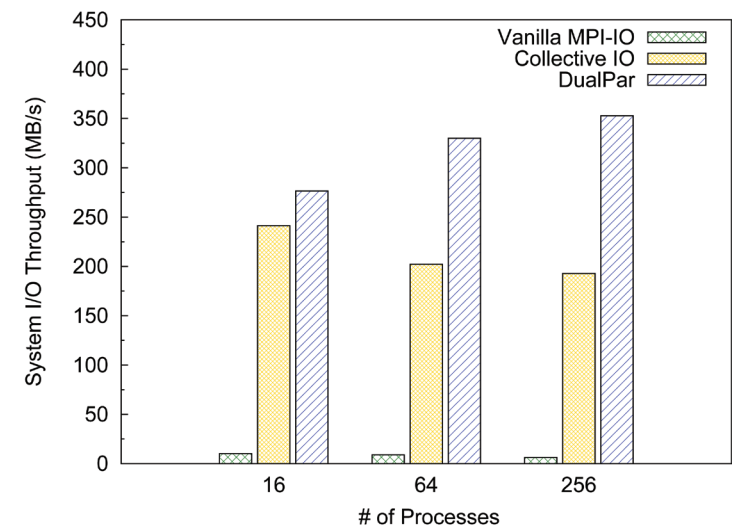
Fig. 2. System I/O throughput with three concurrent instances of BTIO benchmark. We compare throughput with DualPar to those with vanilla MPI-IO and collective IO, as we increase process parallelism from 16, 64, to 256.

[1] Zhang, X. et al., "iTransformer: Using SSD to Improve Disk Scheduling for High-performance I/O," *26th IEEE Int Parallel Distr Process Symp,* to appear (2012).

Optimization Principles for Hardware/Software Co-Design with Applications in Molecular Dynamics

Stephan J. Eidenbenz, CCS-3; Kei Davis, CCS-7;
Arthur F. Voter, T-1; Hristo N. Djidjev, Leonid Gurvits,
CCS-3; Christoph Junghans, T-1;
Susan M. Mniszewski, CCS-3; Danny Perez, T-1;
Nandakishore Santhi, Sunil Thulasidasan, CCS-3

We are developing a framework for hardware-software co-design as a formally posed optimization problem. While the optimization framework will be applicable to multiple problem domains, our target application is molecular dynamics (MD), an exemplar for the need for both weak (larger problems) and strong (faster time to solution) computational scaling, and archetypal of the obstacles thereof. We view co-design as search and selection from a vast space of hardware and software designs that map to performance metrics such as run time (or computational rate), problem size, simulated time duration, energy use, and hardware cost. Successful implementation will provide the unique capability of co-designing hardware and software on a sound formal basis. It will enable effective implementation of physics simulations on future novel computer architectures, and provide a basis for more decisive influence on hardware design. For our MD application domain, we will define plausible regimes that would enable us to study a wealth of poorly understood phenomena of critical importance to the LANL mission.

Hardware-software co-design, perceived as a prerequisite to achieving effective exascale computing, needs to be put on a sound scientific basis such that design decisions for both hardware and software are not made based on colloquial heuristic insights, but rather follow an established scientific procedure by a sufficiently thorough search of realizable hardware and software options.

Figure 1 illustrates our approach, where the hardware (left) and software (right) design boxes represent the spaces of hardware and software solutions, whose combinations lead to performance predictions, the results of which in turn guide an optimization method towards new hardware/software solutions to be tested. The resulting iterative process can be analyzed in formal and informal settings, thus opening doors to established optimization and analysis techniques, while at the same time incorporating sometimes superior but informal human ingenuity. The key research efforts in this approach are:

- efficient enumeration strategies for both the hardware and software design spaces ("What are our options?") that accurately reflect physical constraints and trade-offs in realizable hardware and software designs;

- a multi-scale approach to performance prediction modeling ("How fast will it run?"), where we use cycle-accurate virtual machine emulation, discrete event simulation, graph mapping, and constraint programming as different prediction methodologies; and

- optimization methods to search the design spaces ("How do we select solutions?"), with fast identification of new HW/SW pairs to be tested with the more detailed performance prediction methods.

We define hardware and software designs in a hierarchical fashion. Enumeration of software designs will initially be via correctness-preserving transformations. The enumeration of hardware architectures is done implicitly through parameter spaces and optimization search methods as well as design feasibility checking (human, possibly machine-aided). Our performance prediction methods model at different levels of detail, thus spanning the trade-off between accuracy and scale in both time and size. The performance prediction techniques we will employ include constraint mathematical programming, discrete event simulation, and virtual machine emulation. Optimization heuristics are necessary for the more detailed prediction methods of virtual machine emulation and discrete event simulation. Our framework allows for the adaptation of standard meta-heuristics such as simulated annealing, taboo search, gradient search, and genetic algorithms. The more coarse-grained prediction methods of graph mapping and constraint

Fig. 1. The co-design optimization loop: the optimization method repeatedly applies the performance prediction method to the previously selected hardware/software pair, then selects a new pair based on that predicted performance.

Dashed and dash-dotted lines correspond to the theoretical limit of a perfectly-scaling MD code at peta- and exa-scale, respectively.

Imperfect scaling shows an opportunity for both Co-Design and discovery at current scale.
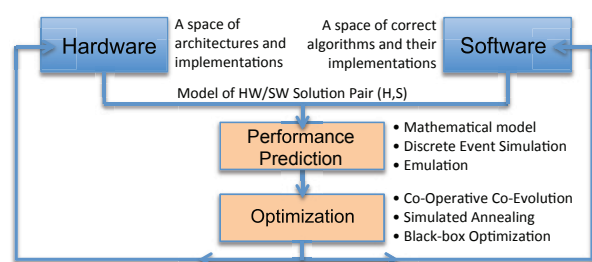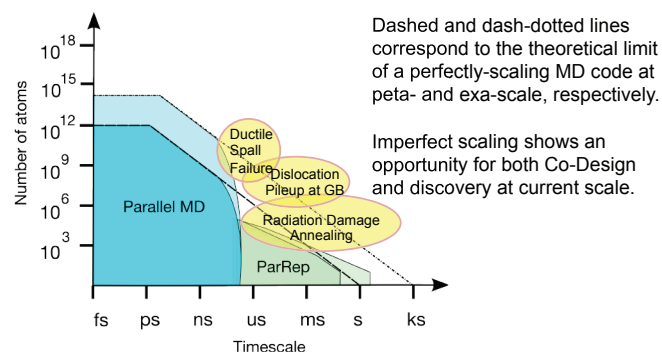
Fig. 2. The need for Co-design to overcome current performance boundaries of MD/AMD. Strongly colored regions correspond to the accessible problem space on a current petascale computer (24 hours on Roadrunner) while paler regions are for an exascale machine. The dashed and dash-dotted lines correspond to the theoretical limit of a perfectly scaling MD code at the peta- and exascale, respectively. Note the capability hole at 10^6–10^9 atoms for microseconds to milliseconds for both peta- and exascale.

programming will optimize through graph algorithms and mathematical programming techniques.

On the application side, our goal is to enable the atomistic simulation tools that can probe the physics of processes such as ductile spall failure under shock conditions, and the evolution of radiation damage. Achieving this requires two orders of magnitude increase in simulated time over current state-of-the-art petascale computing, but direct implementation on an exascale platform would not achieve it. These two material-science problems are of great intrinsic interest to LANL, addressing the response of materials in extreme conditions, and enabling the design of more durable and safe fission power plants.

The severity of the scaling problem becomes obvious when we estimate the behavior of MD or accelerated molecular dynamics (AMD) techniques [1-4] such as parallel-replica methods (ParRep [2]) on an anticipated realization (say in 2020) of an exascale machine with 10^8 processing cores (Fig. 2) [5]. The increased memory will allow simulations up to ~10^{14} atoms for picosecond time scales, and ParRep for very small systems might reach 1-second time scales if the events were very infrequent (once per 10 nanoseconds for 1000 atoms), but essentially no new territory is covered in the middle range. The important regime of 10^6 to 10^9 atoms for microseconds to milliseconds is still totally inaccessible, meaning we will have no direct way to simulate processes such as ductile spall failure, dislocation pileup and release at grain boundaries, or radiation damage annealing involving multiple cascades, all of which are important materials science problems and critical to the Department of Energy mission. In short, if we do not find a new paradigm for implementing MD methods on advanced computer architectures, critical problems in materials science that could in principle be addressed in a powerful way with MD or AMD will remain utterly out of reach for the indefinite future. We believe the answer to

this problem lies in the development of a formal co-design capability that optimizes hardware and software simultaneously.

At the highest level this work will yield a sound basis for conducting disciplined hardware/software co-design as a formally-cast optimization problem. Within this conceptual framework we will provide a number of techniques, at various levels of abstraction, for modeling hardware and its performance. While these may not be fundamentally new, what will be new is their being posed such that they provide a well-defined and consistent interface to an optimizer. We will demonstrate the practical utility of using formal optimization techniques to rapidly identify high value points in the design space. While MD calculations are the focus, the framework and techniques are general.

A successful co-design optimization framework will fundamentally change the way scientific applications are constructed: hardware/software co-design fundamentals will become an essential part of software engineering. The optimization framework will have manual and automated aspects, similar to compiler optimization techniques used in classical compilers for the automated aspects; it will achieve peak performance when the human designer allows the optimizer to test a large set of potential solutions in a high-throughput optimization loop.

[1] Voter, A.F., *Phys Rev Lett* **78,** 3908 (1997).
[2] Voter, A.F., *Phys Rev B* **57,** R13985 (1998).
[3] Sorensen, M.R. and A.F. Voter, *J Chem Phys* **112,** 9599 (2000).
[4] Perez, D. et al., *Annu Rep Comput Chem* **5,** 79 (2009).
[5] Perez, D. et al., *ADTSC Science Highlights 2010,* 130, Los Alamos National Laboratory (2010).

Parallel Log Structured File System Update

Meghan (Wingate) McClelland, HPC-5;
Gary A. Grider, HPC-DO; Adam Manzanares, HPC-5;
John Bent, EMC Corporation

To improve the checkpoint bandwidth of critical applications at LANL we developed the Parallel Log Structured File System (PLFS)[1]. PLFS is a transformative I/O middleware layer placed within our software stack that transparently rearranges a challenging workload, a concurrently written single shared file, into an optimized workload, non concurrently written non-shared component pieces. This reorganized I/O has made write size a non-issue and improved checkpoint performance by orders of magnitude measured to be as much as 150x with improvements in write, read, and meta data performance of our I/O workloads. Under a CRADA with EMC Corporation, LANL and EMC have successfully demonstrated a prototype "burst buffer" and are working together to further enhance, design, build, test, and deploy PLFS. Future work for PLFS includes integration with the Scalable Checkpoint/Restart (SCR) Library, further improving metadata rates, and capturing semantic information about the data streams.

Fig. 1. Index Aggregation Techniques. The figures below represent the workloads generated on the parallel file system and cluster interconnect by PLFS using the original design and our three solutions to improve the read bandwidth of PLFS.
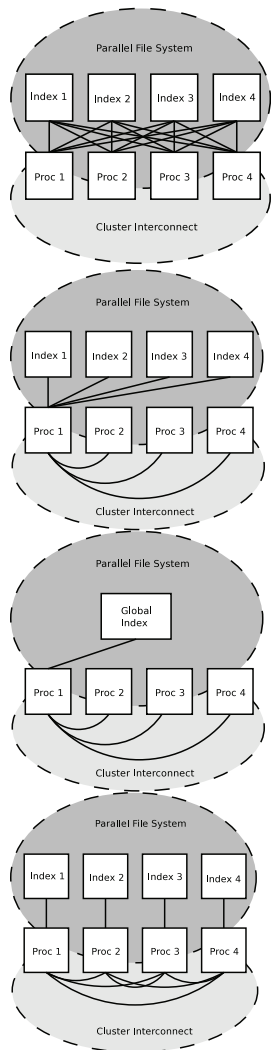
The original architecture of the Parallel Log Structured File System (PLFS) was heavily write optimized and limited read performance. We quickly noticed that with increasing process counts our effective read bandwidth was suffering due to read open times so we developed several techniques to address this problem (Fig 1), which are to aggregate the global index on

- read open with one process and broadcast this result to every other process (Index Broadcast);

- write close, and on read open broadcast the results of the aggregation from one process to all processes (Index Flatten); and

- read open leveraging all processes (Parallel Index Read).

All three share the common property of reducing the number of input/output (I/O) operations that must be conducted by the parallel file system. During a checkpoint read the compute and high speed interconnect resources of cluster machines are largely idle. Our solutions leverage both of these resources and to some degree reduce the amount of concurrent access to files as compared to the original design of PLFS. Bandwidth improvements and faster open and close times are realized with all three methods (Fig. 2).

Metadata management is a difficult challenge for many current parallel file systems. As we move into exascale-class computing, compute systems will have increased component counts which will increase metadata workloads, exacerbating the metadata challenges that we currently face. The flexibility granted by the PLFS layer allows us to alleviate metadata challenges by spreading the I/O workload over multiple physical locations (hashing), turning independent metadata servers into a federated system of servers which improces metadata performance (Fig. 3), and spreading workloads over multiple directories solving the problem of massive-scale file creation in single directories. This increased metadata performance has been demonstrated across two classes of workloads that together represent the checkpoint point I/O patterns of a large class of applications.

The promising effectiveness and flexibility of PLFS has made it necessary to develop PLFS into production software; a regression test suite and quality assurance (QA) procedures have led to improvements in robustness, performance (including archiving and restarting with varying number of processors), and new features. PLFS is in the process of being deployed on current production systems and our plans for exascale I/O include the use of PLFS.

Recently, LANL announced the signing of a new umbrella CRADA with EMC Corporation. As part of this CRADA, LANL and EMC are working together to enhance, design, build, test, and deploy PLFS. This collaboration is focused on support for the Department of Energy's exascale initiative and other data-intensive programs, and is aimed at boosting high-performance computing capability to ensure efficient

(a) Read Open + Write Close Times

(b) Read Bandwidth

Fig. 2. The graph on the left illustrates the time taken to aggregate the global index of a PLFS file with lower times representing better performance. Note: the write close times are included in the graph because the Index Flatten technique aggregates the global index 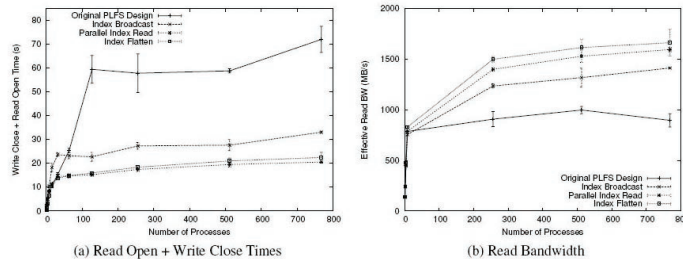on the close of a newly written file. The graph on the right illustrates the effective read bandwidth of the original PLFS design and our read optimizations.

Fig. 3. Graphs of results from a hashed metadata study. In most cases, multiple metadata servers (hashing) dramatically improve file open time. Note that smaller open times are better, so lower lines are better performing for these graphs.

Hashed Metadata (Cores/Volumes Study)

(a) N-1 Write

(b) N-1 Read

(c) N-N Write

(d) N-N Read
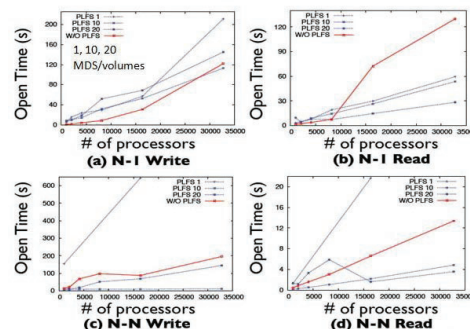
resource utilization on the largest supercomputers in the world.

The first deliverable for this CRADA was the demonstration at the 2011 Supercomputing Conference of a prototype "burst-buffer" storage stack (Fig. 4). Economic projections into the exascale era dictate that the storage stack be re-engineered to incorporate an intermediary layer between the compute nodes and the disk-based scratch storage space. This prototype burst-buffer system was a surprisingly effective exploration within this new architectural space. By accomplishing a complicated integration of both new hardware and software elements in a very short timeframe, LANL and EMC demonstrated their ability to work together closely and effectively.

Using a wind-turbine simulation and visualization workload, the demonstration showed much faster data dump times as well as the ability to co-process the visualization with the simulation, as opposed to the current post-processing model. The very encouraging performance results of this demonstration were of great interest to the community and necessitate more investigation of the burst-buffer concept for exascale computing.

The PLFS project is now looking to further improve metadata rates. Large metadata workloads are created both by user N-N workloads and also by PLFS as it transforms user N-1 workloads. For an exascale system, metadata ingest rates will become a horrible bottleneck as N approaches one billion. Current efforts are ongoing to aggregate multiple logical users' files into a smaller number of physical files to reduce the metadata burden. In addition, the internal PLFS metadata architecture is currently being re-engineered for exascale workloads. One method currently being explored is to use the speed of the compute interconnect workload to shuffle small amounts of data to transform per-process heterogeneous workloads into homogenous ones. This regularity will allow a huge reduction of PLFS's internal metadata.

As part of a tri-lab effort, PLFS is currently being integrated with Scalable Checkpoint/Restart (SCR) [2], a checkpoint in-memory system for N-N codes developed at LLNL.

Finally, the most ambitious and important PLFS project currently ongoing is to capture semantic information about the data streams. This will allow PLFS to know not just where data is stored but what data is stored where. With this knowledge, analysis and visualization workloads can query high-level semantic data knowledge and avoid costly data filtering which can reduce the amount of data read by orders of magnitude. This outcome is related to active storage projects in the past and can enable scientific workloads to benefit from programming models similar to map-reduce.

PLFS is open source software available from
http://sourceforge.net/projects/plfs

Lustre only: post-processing

Lustre with ssdn: co-processing
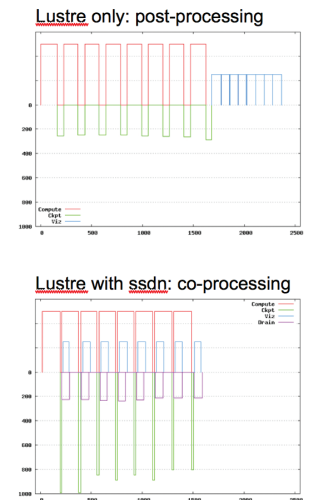
Fig. 4. Graphs of prototype "burst-buffer" demonstration at the 2011 Supercomputing Conference show much faster data dump times as well as the ability to co-process the visualization with the simulation (bottom graph), as opposed to the current post-processing model (top graph).

[1] Bent, J. et al., *Proc Conf HPC Networking Storage Anal,* Article 21; DOI=10.1145/1654059.1654081 http://doi.acm org/10.1145/1654059.16540812009 (2009).
[2] Moody, A. et al., *Proc 2010 ACM/IEEE Int Conf HPC Networking Storage Anal,* **1** (2010).

Randomized Selection on the GPU

Laura M. Monroe, HPC-5; Joanne R. Wendelberger,
Sarah E. Michalak, CCS-6

We discuss the implementation of a fast, memory-sparing probabilistic top **k** selection algorithm on the graphics processing unit (GPU) [1]. This probabilistic approach permits exploitation of the massive parallelism of the GPU. This algorithm is linear in list length and is much faster than previously known GPU selections.

Fig. 1. Flowchart representing the selection algorithm. Those kernels making heavy use of the GPU are shaded, with the darkest being the most time-consuming.
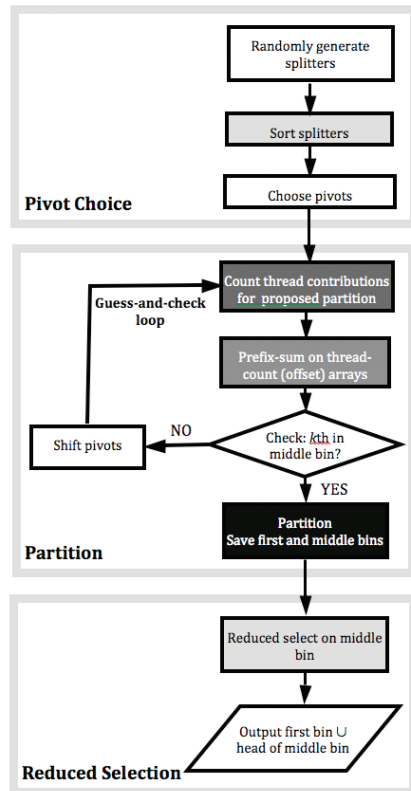
Selection of the top k elements from a strictly weakly ordered set is one of the classic computer science algorithms [2]. Selection has application to a wide variety of statistical, computational, and database-processing problems, such as computer vision, robotics, data mining and image processing. Our initial motivation was in support of a graphics processing unit (GPU) implementation of the CLEAN algorithm, used in radioastronomy to remove noise from images that are generated from multiple antennas [3,4].

Las Vegas algorithms like this selection are a form of stochastic optimization, and optimize time spent and memory used. They can be well suited to more general parallel processors with limited amounts of fast memory. For problems having a quantifiable and efficiently checkable randomized way to guess an answer, the Las Vegas approach may be much more efficient than a direct calculation.

Selection of GPU. The GPU is an accelerator that is seeing increasing use in high-performance computing. It has a massively parallel single instruction, multiple data (SIMD) architecture and a hierarchical memory structure. This architecture is best suited to parallel algorithms that make use of the many processors without demanding much inter-processor data movement. Advantages of the GPU include massive parallelism providing much performance improvement (for parallel algorithms). Disadvantages include the small amount of fast memory close to each processor, necessitating careful and limited data placement, and the constrained bandwidth from the central processing unit (CPU) to GPU.

The naïve approaches to selection do not map well to the GPU. These compare elements on a global basis, so they call for an ample amount of fast memory to hold elements for comparison, or else a large amount of data movement across memory hierarchies. Selection via sort is another approach: after a sort, the top k elements are trivially extracted. Although fast sorts exist on the GPU, this method does more work than needed, and therefore is wasteful. If the kth element were only known ahead of time, selection would be well-suited to the GPU—elements could be read into the processors, compared against the known kth key, and saved if their rank is higher than the kth. This process is highly parallel and requires little inter-processor data movement. Unfortunately, neither the kth element nor the list distribution is known before execution.

It is possible, however, to guess the kth element probabilistically, and thus proceed as if the kth were known. Two list elements are found between which the kth element is located with arbitrary probability. This probabilistic guess eliminates the need for global data availability and transforms the problem into the highly parallel version described above. Although this algorithm is probabilistic, it always calculates a correct, unordered set of the top k elements, always terminates, and terminates after only one iteration with arbitrarily chosen probability.

This randomized selection is an example of a Las Vegas algorithm [5]. A Las Vegas algorithm is one of a class of algorithms for which the methods are probabilistic, but the result is always correct. Similar selections were developed by Motwani and Raghavan [6] and by Bader [7], but these methods used a different method of guessing the kth element.

Fig. 2. Timings for different k/n, where k is the rank of the element selected for and n is the list length. Each line marked by a percentage represents a different value for k/n. These are compared to three other methods of GPU selection.

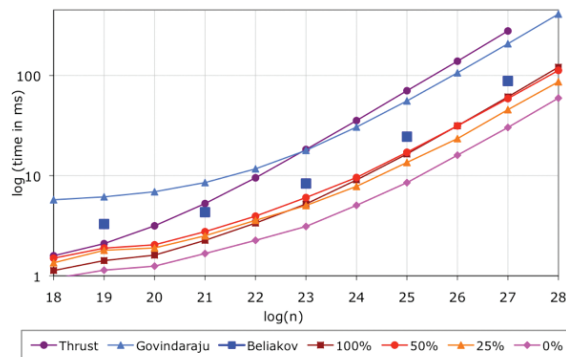*Randomized Selection Algorithm.* The algorithm proceeds via an iterative, probabilistic guess-and-check process for a three-way partition. The algorithm is composed of three parts: 1) the probabilistic guess of two pivots between which the kth element is likely to fall, 2) partition based on the pivots, and 3) reduced selection on the smallest (middle) part of the partition, holding the kth element. Both keys and values are extracted. The original list is untouched.

To guess the two pivots, we use a set of randomly chosen keys (called splitters) from the list. We sort the splitters and use them to conceptually partition the list, thus "flattening out" the distribution. This conceptual partition produces buckets that typically contain roughly the same number of keys, so one can reason probabilistically about them. We use this conceptual structure to estimate the probability of the kth key falling into each bucket.

This estimate makes use of a binomial approximation to determine the probability that the kth key is in a specific bucket or set of buckets. If the kth key lies outside the range of the splitters, the approximation may be poor for the end buckets, but the impact will be minimal if the tail probabilities associated with the end buckets are small. Numerical studies confirm that the method performs well. The bucket probabilities are added and a concatenation of buckets is chosen so that the concatenation 1) contains the kth key with desired probability, and 2) is relatively small in size. The two splitters bounding the concatenation then serve as pivots.

Keys and associated values are partitioned into bins defined by the pivots and counted. If the count shows that the kth key is in the bin bounded by the pivots, the algorithm completes with a selection on this relatively small bin. Otherwise, new pivots are chosen and the process iterates.

Performance on the GPU. This algorithm makes excellent use of the parallelism available on the GPU. Our largest runs used 64 million

threads, delivering good thread-level parallelism. The algorithm is not particularly compute-intensive, but truly benefits from the massive parallelism of the GPU. The latency from the main memory to the GPU across the PCIe bus can be many times that of selection itself, so if the overall calculation is on the GPU, it is far better to efficiently select there as well. The limiting factor on the maximum size of the original list is the amount of global memory on the GPU.

We tested on the NVIDIA Quadros 5000 and 6000, and GeForce GTX 285 graphics cards. We show results from runs on the 6000. We compared randomized selection, selection-via-sort using Thrust radix sort [8], construction of the kth element [9], and construction by minimization of a convex function [10]. We experimented on lists of 32-bit integer keys, of length from 2^{18} to 2^{29}. These were ordered randomly, randomly with high entropy, sorted and sorted backwards, and also included a real 32-bit 2^{18}-pixel image from radio astronomy. Our results show: 1) speedups for larger lists are three to six times faster than selections via sort and direct construction, and one to two times faster than selections via minimization of a convex function; and 2) list sizes can be processed that are up to four times longer than those possible using sort.

[1] Monroe, L. et al., "Randomized Selection on the GPU," *Proc ACM SIGGRAPH Symposium High Perf Graphics 2011* (2011).
[2] Knuth, D., *Sorting by Selection. The Art of Computer Programming,* Vol, **3,** Third Edition, Addison-Wesley (1997).
[3] Högbom, J.A., *Astron Astrophys Suppl* **15,** 417 (1974).
[4] Clark, B.G., *Astron Astrophys* **89,** 377 (1980).
[5] Babai, L., "Monte Carlo Algorithms in Graph Isomorphism Testing," Tech. Rep. DMS 79-10, Universite de Montreal (1979).
[6] Motwani, R. and P. Raghavan, *Randomized Algorithms,* Cambridge University Press (1995).
[7] Bader, D.A., *J Parallel Distr Comput* **64,** 1051 (2004).
[8] Hoberock, J. and N. Bell, http://code.google.com/p/thrust/2010 (2010).
[9] Govindaraju, N. et al., "Fast Computation of Database Operations Using Graphics Processors," *Proc ACM SIGMOD* (2004).
[10] Beliakov, G., arXiv:1104.2732v1 (2011).

Atomic, Nuclear, and High-Energy Physics

In this section we highlight examples of recent research that is representative of work done in ADTSC in support of the Laboratory and DOE mission. The Laboratory has developed a major capability in this area and the articles in this section exemplify the breadth of our work.

This section includes discussions on the interaction of radiation with atoms, quantum computation, quantum molecular dynamics, the behavior of plasmas under nonlocal equilibrium conditions, molecular dynamics simulations of plasmas, quantification of margins and uncertainties, radiative transfer in high energy physics applications, and the implicit Monte Carlo method.

Time-dependent Calculations of Electron Energy Distribution Functions for Cold Argon Gas in the Presence of Intense Black-Body Radiation

Joseph Abdallah Jr., James P. Colgan, T-1

Boltzmann electron kinetic simulations are performed to study the time development of the electron energy distribution in a plasma that results from a cold argon gas subject to a black-body radiation source (100–300 eV). The study provides insight into the role of ionized electrons on the kinetics during a period of irradiation. The simulations are performed without any assumptions of electron temperature. The distributions are calculated as a function of time through 1 ps using Boltzmann kinetics, including the appropriate processes that alter state populations and electron energy. The processes included in the electron and atomic kinetics are collisional excitation/de-excitation, photo-excitation/decay, photo-ionization/radiative-recombination, collisional ionization, and auto-ionization/di-electronic capture. In addition, terms are included in the electron kinetics to account for electron-electron interactions and for free-free radiation absorption and emission. Results are presented that follow the evolution of the ionization state effective temperature as well as the electron energy distribution function (EEDF). The role of inelastic electron collisions, photo-ionization, auto-ionization, resonant radiative excitation, and electron-electron interactions are discussed.

Fig. 1. The black-body radiation field ($cm^{-3} eV^{-1}$) for radiation temperatures of 100, 200, and 300 eV.

Fig. 2. The calculated average charge state (Z^*) as a function of time (picoseconds) for radiation temperatures of 100, 200, and 300 eV. The solid lines are the results of the full model while the dashed lines correspond to calculations where collisional excitation and ionization have been neglected.
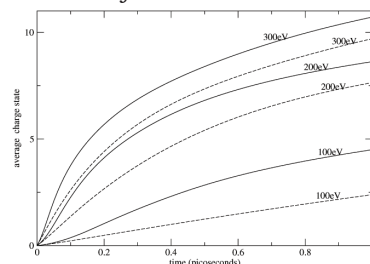
High-energy density physics devices such as the Z pinch [1,2] have been used to create astrophysical laboratory plasmas by irradiation of matter with an intense source of X-rays. The emerging X-ray free electron laser (XFEL) [3] technology also has many interesting applications involving laser-matter interactions (for example, [4]). The Z pinch is capable of generating black-body radiation of several hundred eV, while XFEL produces a monochromatic X-ray source at tunable wavelengths. Each is capable of providing radiation capable of exciting and ionizing outer- and inner-shell electrons of atoms through direct and Auger processes. The free electrons then form a plasma whose properties are defined by their electron energy distribution function (EEDF). Both photon and electron processes occur to varying degrees after the initial plasma is formed while the radiation source is active. Spectroscopic properties of plasmas can be very sensitive [5-7] to the EEDF, especially the high-energy behavior.

This paper presents proof of principle simulations of the plasma EEDF, up to 1 ps, when argon gas is irradiated by black-body (100-300 eV) radiation, and provides insight about the important processes occurring in the plasma under these conditions. Unfortunately, there are no measurements of the EEDF for comparison purposes. The authors are unaware of any similar calculations under these conditions. The approach used is an extension of previous calculations [8,9] used to model the X-ray spectra resulting from the interaction of a high-power ultra-fast laser with argon clusters. In the latter work, an EEDF calculated from a particle in cell (PIC) code was used as an initial condition. This EEDF was propagated in time by the simultaneous solution of atomic rate equations and Boltzmann electron kinetics. The atomic kinetics keeps track of atomic state populations while the electron kinetics keeps track of the distribution of electron energy. The two, are coupled to each other. Electron impact collisional excitation, de-excitation, ionization, and electron-electron collisions were used in the electron kinetics.

We start out with cold neutral argon gas and follow the EEDF directly, again using both atomic and electron kinetics. The electron kinetics capability was enhanced for this problem. Photo-ionization, radiative recombination, auto-ionization, di-electronic capture and free-free absorption and emission were incorporated in the electron kinetics package. A reduced detailed configuration accounting (RDCA) atomic model [10,11] was used. RDCA is an average of more detailed atomic models designed to be much faster while approximately preserving the atomic kinetics.

For the model calculations considered here, the atom number density was taken to be 1×10^{20} cm^{-3}, and initially all the ions were assumed to be in the ground state of the neutral argon. One hundred time intervals between 0 and 1 ps were chosen and the scheme described above was used to obtain the EEDF and state populations as a function of time. The external radiation field $G(h\nu)$ remained constant for the entire time period and temperatures of 100 eV, 200 eV and 300 eV were used in the simulations. The radiation field corresponding to these temperatures
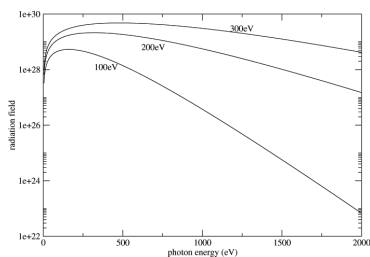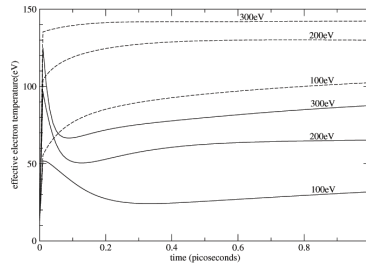
Fig. 3. The calculated effective temperature (kT*) in eV as a function of time (picoseconds) for radiation temperatures of 100, 200, and 300 eV. The solid lines are the results for the full calculation while the dashed lines correspond to calculations where collisional excitation and ionization have been neglected.
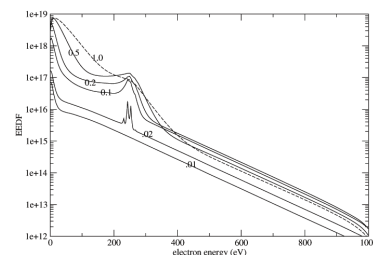

Fig. 4. The calculated electron energy distribution function (EEDF) in cm⁻³ eV¹ as a function of electron energy for 0.01, 0.02, 0.1, 0.2, 0.5, and 1.0 ps during irradiation by a 100-eV black-body radiation field.

is shown in Fig. 1. The EEDF was divided into 500 equally spaced bins from 0 to 1 KeV. The average charge state (Z^*) and effective temperature (kT^*) of electrons are calculated from the EEDF as a function of time. Note that after the time and energy grids have been carefully chosen, the only input to the model is the initial conditions and the radiation temperature.

The average charge state (Z^*) is plotted (solid curves) as a function of time for radiation temperatures of 100, 200, and 300 eV in Fig. 2. Note that the degree of ionization increases with increasing temperature. For 200 and 300 eV, the ionization increases rapidly at first and then levels out, while the ionization is much slower at 100 eV. The predicted Z^* at 1 ps is 4.5, 8.6, and 10.7 for 100, 200, and 300 eV, respectively.

Figure 3 shows the calculated effective electron temperatures as a function of time for the same three cases. Each case shows an initial rapid heating of electrons, followed by rapid cooling until about 0.2 ps, and then slow heating. The initial rapid heating is caused by ionization in the neutral gas due to the radiation field. After sufficient free electrons are created, the plasma cools due to inelastic collisions of electrons with atoms and ions through excitation and ionization processes. The cooling occurs much faster at 300 eV then 100 eV because the higher rate of ionization results in a larger free electron density at 300 eV. After the plasma has cooled sufficiently, the electrons are then again slowly pumped by the radiation source. The dashed curves in Fig. 3 have been included in order to illustrate the effect of cooling. These curves correspond to the same radiation temperatures except that the effects of electron collisional excitation and ionization have been omitted. Here the effective temperature rises rapidly for each case and then levels off. This cooling effect at a radiation temperature of 300 eV accounts more than a 55-eV difference in the electron temperature.

It is interesting to compare the influence of ionization due to electrons versus the radiation field. The dashed curves in Fig. 2 correspond to calculations where collisional excitation and ionization have been neglected in both the electron and atomic kinetics. At a radiation temperature of 100 eV, about half the ionization is due to electrons. Note that as the temperature of the radiation field increases the role of the electrons diminishes as one would expect. At a radiation temperature of

300 eV and 1 ps the electrons account for a small yet significant amount of ionization, about one charge state in ten.

Figure 4 shows the calculated EEDF at various times for the 100 eV radiation temperature case. Note the EEDF does not assume a Maxwellian form over the entire time interval and electron energy range. At .01 ps the EEDF is formed by direct photo-ionization and has a tail (high energy part) that falls like its cross-section. The low photon energy behavior is produced by a combination of the peak of the photo-ionization cross-section and the electron-electron interaction, which tends to smooth and equilibrate the electron distribution. At .02 ps a bump appears at electron energies above 200 eV. This bump corresponds to the production of Auger electrons formed by inner shell excitations and ionizations from the n=2 (L) shell of argon by the radiation field. Note that as time increases more free electrons are produced, the electron-electron interactions become stronger, and the bump gets smoother, but still persists even at 1 ps. At energies below 100 eV for later times, the Maxwellian shape takes form.

Note that EEDF calculations also include the production Auger electrons from K-shell excitation and ionization. However, K-shell vacancies have little effect on the results because fewer photons have energies in excess of 3000 eV, which is required for exciting such states. No K-shell bumps appear in the EEDF since they occur beyond the electron energy grid used here.

[1] Mancini, R.C. et al., *Phys Plasma* **16,** 041001 (2009).
[2] Hall I.M. et al., *Astrophys Space Sci* **322,** 117 (2009).
[3] Pelliegrini, C., *Proc 16th Topical Conf Atomic Process Plasma,* 113 (2009).
[4] Rosmej, F.B. et al., *High Energy Density Physics* **3,** 218 (2007).
[5] Hansen, S.B. and A.S. Shlyaptseva, *Proc 14th Topical Conf Atomic Process Plasma,* 213 (2004).
[6] Abdallah, J. Jr. et al., *J Quant Spectros Radiative Tran* **62,** 1 (1999).
[7] Abdallah, J. Jr. et al., *Phys Scripta* **53,** 705 (1996).
[8] Abdallah, J. Jr. et al., *Phys Rev A* **68,** 063201 (2003).
[9] Sherrill, M.E. et al., *Phys Rev E* **73,** 066404 (2006).
[10] Abdallah, J. Jr. and M.E. Sherrill, *High Energ Density Phys* **4,** 124 (2008).
[11] Abdallah, J. Jr. et al., *High Energ Density Phys* **5,** 204 (2009).

Non-Demolition Adiabatic Measurement of the Phase Qubit State

Gennady P. Berman, T-4; Alexander A. Chumak,
Institute of Physics, Kiev; Dimitry I. Kamenev, T-4;
Darin Kinion, LLNL; Vladimir I. Tsifrinovich,
Polytechnic Institute at New York University

An adiabatic method for a single-shot non-demolition measurement of the phase qubit is suggested [1]. The qubit is inductively coupled to a low-frequency resonator, which in turn is connected with a classical measurement device (phase meter). The resonator drives adiabatic oscillations of the supercurrent in the qubit loop. The back reaction of the qubit loop on the resonator depends on the qubit state. By measuring the phase shift of the resonator's oscillations one can determine the state of the qubit. Numerical computations with available experimental parameters show that the phase difference between the two qubit states increases at a rate of 0.0044 rad/ns with a fidelity of about 0.9989 and a measurement time of about 100 ns. The fidelity of the measurement is estimated by taking into consideration possible quantum transitions inside and outside of the qubit manifold. An increase of the phase difference is possible, but it is linked to a reduction of the fidelity. The requirements for the reproducibility of the qubit and resonator parameters are formulated.

The idea of our approach is the following. We assume that the "qubit loop" (QL) is inductively coupled to the superconducting low-frequency measurement resonator (MR). There is an "adiabatic switch" that allows one to "turn on" or "turn off" the QL-MR coupling adiabatically with respect to the QL and "instantaneously" with respect to the MR. The adiabatic switch can be implemented by a variable mutual inductance, which is already developed. In this design, a current bias applied to a direct current superconducting quantum interference device (DC SQUID) controls the screening current, which influences the inductive coupling between two circuits.

The scheme of our design is shown in Fig. 1. The supercurrent in the MR oscillates with a much lower frequency than the frequency of the phase qubit. The MR oscillations cause oscillations of the flux in the QL. As a result, the positions of the minima of the potential energy of the qubit adiabatically oscillate. Thus, the supercurrent in the QL adiabatically oscillates with the frequency of the MR. The back action of the QL oscillations on the MR causes a phase shift in the MR

oscillations. This phase shift depends on the qubit state. By measuring the phase shift of the MR oscillations one can determine the state of the phase qubit. Note that the qubit, which is placed initially in one of its basis computational states (ground or excited), remains in the same state during the measurement. If the qubit is placed initially into a superposition of the two basis states, it is expected to collapse quickly to one of its basis states. This should happen because the z-component of the phase qubit Bloch vector is an adiabatic invariant of the QL-MR dynamics, which does not change in the process of the phase measurement. Thus, our scheme describes a non-demolition projective measurement of the phase qubit. Note that the interaction between the QL and MR in our scheme is supposed to be strong enough so that the MR phase shift is not negligible in spite of the large difference between the qubit and MR frequencies. The low frequency oscillations in the resonator are supposed to be amplified with a microstrip SQUID amplifier (MSA) developed at LLNL, with almost quantum-limited noise. The phase of the amplified oscillations is measured with a phase meter.

Fig. 1. The measurement scheme suggested in our work. The superconducting loop interrupted by the Josephson junction (JJ) is coupled inductively to a low-frequency measurement resonator (MR). M is the controllable mutual inductance, which can be turned on; ϕ_p is the external permanent flux which biases the QL; L_p and C_r are, respectively, the inductance and the capacitance of the MR; L is the QL inductance.
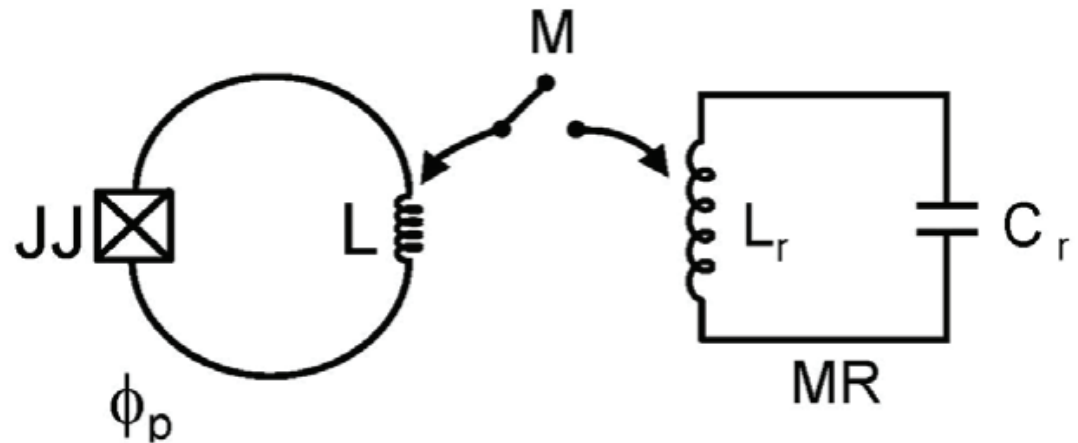
[1] Berman, G.P. et al., *Quant Inform Comput* **11,** 1045 (2011).

Energy-conserving, Linear-scaling Quantum Molecular Dynamics

Marc J. Cawkwell, Anders M. N. Niklasson, T-1

Molecular dynamics (MD) simulations are used heavily in materials science, chemistry, and biology to study the evolution of structures, defects, and non-equilibrium phenomena at the atomic scale. In an MD simulation atoms move over a number of finite time steps according to the force acting on them. These forces are computed from the interatomic potential that gives the potential energy of the system as a function of the relative positions of all of the atoms. The ability of a simulation to capture the system of interest with high fidelity is determined almost entirely by the physical accuracy of the interatomic potential. It is well established that explicitly quantum mechanical models provide the most accurate descriptions of bonding, but it is not possible to employ these methods in large-scale simulations owing to their prohibitive computational cost.

The scope of quantum-based molecular dynamics (MD) simulations has been limited primarily by the cubic scaling, $O(N^3)$, of the computational cost of solving the quantum eigenvalue problem on the number of atoms, N [1]. Simulations that involve more than a few hundred atoms require a computational cost that scales linearly, $O(N)$, with the number of atoms. Methods have been developed that enable the $O(N)$ scaling of quantum-based potentials that are based on, for example, splitting a large system into a number of non- or weakly interacting smaller ones [1]. The fastest $O(N)$ methods discard long range interactions entirely and enforce "nearsightedness." While the latter approach is physically intuitive, it leads to errors in the electronic structure and forces that may corrupt an MD trajectory such that the total energy is not conserved. It was not certain that $O(N)$ force computations and energy-conserving MD could ever be reconciled-
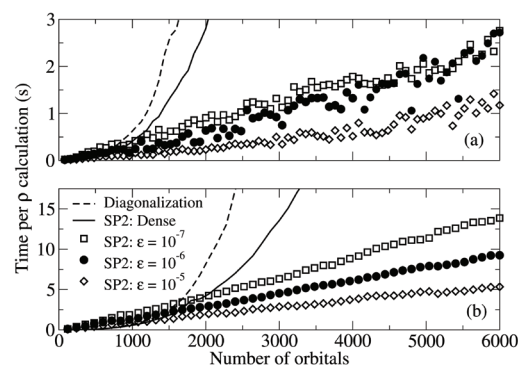
We have developed the capability for performing energy-conserving quantum MD with an $O(N)$ scaling of the computational cost in the LANL-developed code LATTE [2]. LATTE is based on a self-consistent tight-binding model [3], which is the simplest quantum scheme that captures all of the most important interactions in organic molecules–the formation of covalent bonds, charge transfer between species of differing electronegativity, and long range electrostatics. The computation of the energy and forces requires the solution of a Schrödinger-like equation for the density matrix. The computation

of the density matrix is by far the slowest step in the calculation and, furthermore, when traditional approaches are used it as an undesirable $O(N^3)$ scaling. We achieved $O(N)$ scaling for this step by implementing Niklasson's second order spectral projection purification (SP2) algorithm [4] in sparse matrix algebra where only non-zero matrix elements are stored and operated on. By discarding elements whose absolute value is smaller than a threshold, ε, at each iteration in the SP2 algorithm, we maintain very high levels of sparsity and enforce nearsightedness. In Fig. 1 we plot the time required to compute the density matrix, ρ, as a function of system size for liquid methane and polyethylene using standard diagonalization techniques, the SP2 algorithm in dense-matrix algebra, and the $O(N)$ implementation of the SP2 algorithm [5]. It is evident that the $O(N)$ scheme becomes rapidly orders of magnitude faster than the $O(N^3)$ methods as the number of atoms increases.

The analysis of the density matrices computed using the $O(N)$ schemes reveals errors that increase as ε increases. However, even at the coarsest threshold of $\varepsilon = 10^{-5}$ we observe an average error in the interatomic forces of less than 0.02% per atom [5]. Nevertheless, errors of this magnitude are large compared with the numerical precision.

During Born-Oppenheimer MD the atoms following classical trajectories are computed from forces that are evaluated at the self-consistent electronic ground state. Even in the $O(N^3)$ limit these simulations may show a systematic drift in the total energy when the iterations for self-consistency are not converged fully. It was shown recently that this systematic drift can be removed without resorting to full convergence

Fig. 1. Time per density matrix calculation as a function of the number of orbitals for (a) liquid methane, (b) a polyethylene molecule.
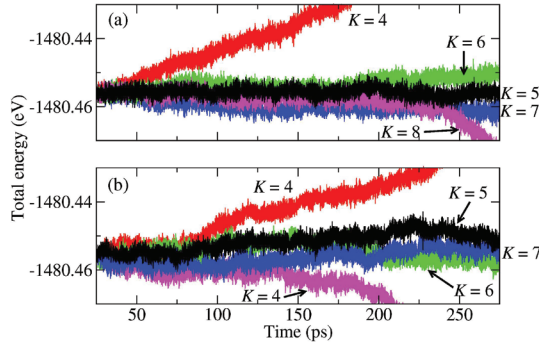
Fig. 2. Total energy versus time for extended Lagrangian Born-Oppenheimer MD simulations with dissipation schemes. (a) Exact forces (b) Approximate forces from O(N) scheme with $\varepsilon = 10^{-5}$.

Fig. 3. Time-averaged distribution functions for gas-phase $C_{100}H_{202}$. The black and red lines correspond to the simulations computed using exact forces and approximate forces computed using the O(N) scheme with $\varepsilon = 10^{5}$, respectively. The plots have been offset for clarity.

by propagating the electronic degrees of freedom with a geometric integrator. The corresponding extended Lagrangian Born-Oppenheimer MD formalism has been shown to yield stable, energy-conserving trajectories in both first principles and self-consistent tight-binding simulations [6].

The propagation of the electronic degrees of freedom in extended Lagrangian Born-Oppenheimer MD is perfectly lossless, such that errors in the computation of the density matrix accumulate and may eventually corrupt the trajectory. As a result, it is necessary to introduce dissipation into the propagation to counteract the accumulation of numerical noise [7]. Dissipation in the propagation of the electronic degrees of freedom, $n=\{n_k\}$, which are the starting guesses for the set of partial charges, is given by a Langevin-like dynamics where the modified Verlet integrator at time step i is:

$$n_{i+1} = 2n_i - n_{i-1} + \kappa(q_i - n_i) + \alpha \sum_{k=0}^{K} c_k n_{i-k} \quad (1)$$

The level of dissipation is controlled by the parameters κ and α, and by the K + 1 coefficients c_κ. In a series of simulations of 50 methane molecules at liquid density and a temperature of 300 K, we investigated the effects of dissipation on the stability of MD trajectories computed with approximate forces obtained from the O(N) scheme with $\varepsilon = 10^{-5}$.

These trajectories, which are plotted in Fig. 2, show that the dissipation scheme K = 5 provides the systematic drift in the total energy of not exceeding 0.05 µeV atom^{-1} ps^{-1} for both the exact and O(N) simulations [5].

The ability of O(N) quantum MD to reproduce the results of otherwise identical simulations performed with exact forces was assessed using a simulation of a $C_{100}H_{202}$ polyethylene molecule in the gas phase. The radius of gyration of the molecule computed over 0.55 ns was 16.2 ± 3.0 Å and 16.6 ± 2.9 Å for the exact and O(N) simulations,

respectively. The average atomic structure of the molecules was assessed using the distribution function:

$$\Gamma_{AB}(R) = \frac{1}{N_A N_B \Delta R} \sum_{\substack{i=1 \\ i \in A}}^{N_A} \sum_{\substack{j=1 \\ j \in B}}^{N_B} \delta(R_{ij}; R + \Delta R), \quad (2)$$

where A and B label species, N_A and N_B are the numbers of atoms of species A and B, R_{ij} is the distance between atoms i and j, R is the radial distance from atom i, and $\delta(R_{ij}; R + \Delta R) = 1$ if $R \leq R_{ij} < R + \Delta R$ and is zero otherwise [5]. The time-averaged distribution functions for carbon-carbon (C-C), carbon-hydrogen (C-H), and hydrogen-hydrogen (H-H) pairs are plotted in Fig. 3. The $\Gamma(R)$ computed from the exact and O(N) forces are indistinguishable, which strongly suggests that the application of a numerical threshold to enforce nearsightedness in the computation of the density matrix does not affect the ability of the model to capture the salient features of bonding in covalent molecules.

In summary, we have demonstrated for the first time that extended Lagrangian Born-Oppenheimer MD with optimal dissipation enables stable, energy-conserving simulations of extended quantum systems with approximate forces and a low pre-factor O(N) cost. Hence, there is no obstacle to the application of explicitly quantum methods to large-scale simulations in materials science and biology in those situations where empirical potentials are known to be inadequate.
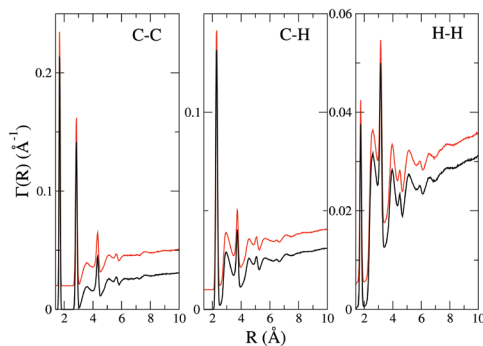
[1] Goedecker, S., *Rev Mod Phys* **71,** 1085 (1999).
[2] Sanville, E.J. et al., "LATTE," LANL Technical Report LA-CC 10-004; http://savannah.nongnu.org/projects/latte (2010).
[3] Elstner, M. et al., *Phys Rev B* **58,** 7260 (1998).
[4] Niklasson, A.M.N., *Phys Rev B* **66,** 155115 (2002).
[5] Cawkwell, M.J. and A.M.N. Niklasson, *Phys Rev Lett,* to appear (2012).
[6] Niklasson, A.M.N. et al., *Phys Rev Lett* **97,** 123001 (2006).
[7] Niklasson, A.M.N. et al., *J Chem Phys* **130,** 214109 (2009).

LANL Contributions to the NLTE-7 Code Comparison Workshop–
Three Test Cases

Gregory S. J. Armstrong, T-1;
Christopher J. Fontes, XCP-5; James P. Colgan, T-1;
Honglin Zhang, XCP-5; David P. Kilcrease,
Joseph Abdallah, Jr., T-1

We report here on three of the LANL contributions to the recent NLTE-7 Code Comparison Workshop that was held in Vienna, Austria, in December 2011. The purpose of the workshop is to provide a detailed comparison among kinetics codes that simulate the properties of nonlocal-thermodynamic-equilibrium (NLTE) plasmas. Comparisons are made for a range of plasma properties such as ion populations and radiative losses. The LANL contributions were made using the LANL suite of atomic physics codes and the ATOMIC kinetics code, and included calculations for all of the test problems. In general, the LANL contri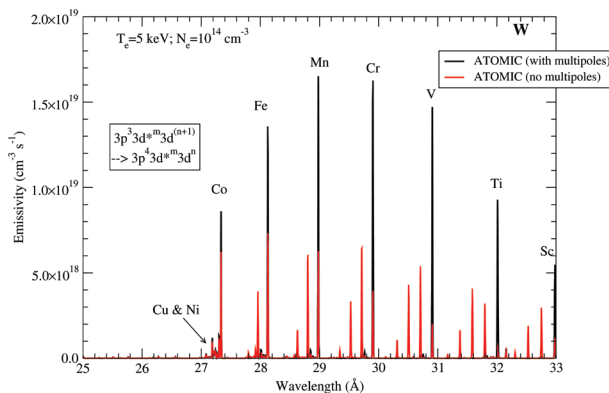butions compared very well to the other submissions, which came from various kinetics codes from institutions worldwide.

The NLTE Code Comparison Workshops [1] were first organized more than 10 years ago in an effort to provide critical comparisons between atomic kinetics codes that model populations of nonlocal-thermodynamic-equilibrium (NLTE) plasmas and their radiative properties under a range of plasma conditions. Such comparisons are particularly useful since there are few (if any) experiments with which to compare such kinetics calculations in large portions of the range of plasma conditions of interest. The Code Comparison Workshops are held every few years, and typically attract from 10 to 15 different code contributions from institutions worldwide. Usually a small number of test cases are chosen, with the intention of making a very detailed comparison among the different calculations at the workshop within a highly focused interactive forum. A dedicated website (hosted by Yu. Ralchenko at the National Institute of Standards and Technology [NIST]) was constructed to facilitate detailed code comparisons. These workshops are very useful to the atomic physics effort at LANL, since they provide important verification and validation of our atomic physics and kinetics modeling codes, which is the principal set of codes used to generate local-thermodynamic-equilibrium (LTE) and NLTE emissivities and opacities.

The seventh workshop, which was held in Vienna, Austria, in December 2011, compared four test cases, each of which offered some insight into plasma properties at various conditions. The LANL contributions to the workshop were made using the LANL suite of atomic physics codes and the ATOMIC kinetics code [2]. The calculations were made in a close collaboration of T-1 personnel (Gregory Armstrong, James Colgan, David Kilcrease, and Joe Abdallah) and XCP-5 staff members (Chris Fontes and Honglin Zhang). Much of the work performed for this workshop is expected to result in journal publications. In this report, we summarize three workshop test cases, with particular emphasis on the LANL calculations made for each case. We also briefly describe the conclusions reached at the workshop.

The first test case considered the properties of neon (Ne) plasma. This case was designed to explore the convergence of kinetics modeling with respect to the highest principal quantum number included within a model. Relatively cold electron temperatures (10 and 30 eV), and a range of electron densities (10^{14}, 10^{18}, and 10^{20} cm^{-3}) were chosen, so that the plasma properties arose from open-shell ions, which are often more difficult to model. (In previous workshops, it was found that when the plasma conditions resulted in plasmas dominated by a closed-shell ion stage, most kinetics codes were in good agreement and so the comparisons were of limited value.) Calculations were then made for six different cases per temperature/density combination, where each of the six calculations used a different maximum principal quantum number in the list of configurations employed. Two further types of LANL calculations were made, one using a configuration-average approximation, and one using a fine-structure approximation. The latter set of calculations included full configuration-interaction between the configurations used, resulting in a set of fine-structure (LSJ) levels. When considering emission spectra, fine-structure calculations are

Fig. 1. Calculated emission spectra for W at an electron temperature of 5 keV and an electron density of 10^{14} cm^{-3}. We compare ATOMIC calculations that include multipole radiative transitions with ATOMIC calculations that omit multipole transitions. The labels on the emission features indicate the ion stage of W from which the transitions arose.

generally more accurate than configuration-average calculations, but fine-structure calculations usually are much more time-consuming and computationally demanding to perform. The emission spectra from the ATOMIC calculations made at the various temperature/density cases resulted in a very large number of bound-bound transitions in almost all cases. Convergence of the emission spectra was obtained for almost all temperature/density cases using a model with a maximum principal quantum number of 12.

The next workshop case considered here was a study of krypton (Kr) plasma. This case was unique due to the availability of an experimental spectrum [3], measured several years ago in France, with which to compare the contributed calculations. These previously published experimental results were initially compared to kinetics calculations with only moderate agreement found between experiment and theory. Two sets of ATOMIC calculations were made for this case. Full fine-structure calculations were performed using a model containing a relatively small number of configurations for the ion stages of interest (centered around Ne-like Kr). A mixed unresolved transition array (UTA) calculation was also performed, in which the population kinetics were calculated within a configuration-average approximation, but where detailed lines were used to accurately model the emission spectrum. The ATOMIC calculations were in reasonable agreement with the measured spectrum. It became apparent that opacity effects modified the measured spectra and had to be taken into account in the calculations. Work is ongoing to improve how ATOMIC treats such opacity effects in the spectral generation.

The final workshop case chosen was an examination of tungsten (W) plasma. Tungsten is an important element in magnetic fusion plasma devices, and will be used extensively in the planned ITER (International Thermonuclear Experimental Reactor) device [4]. It is therefore important to understand the radiative properties of W at conditions expected in magnetic fusion plasmas. This is the third time W has been chosen as an NLTE workshop case. The large number of electrons in W, and the need for relativistic treatments of the atomic structure and collision processes in the W ions of interest complicate comparisons. Calculations were made for a range of electron temperatures from

2.5 keV to 12 keV, at an electron density of 10^{14} cm^{-3}. In Fig. 1 we present an example of the emission spectrum for the 5-keV case. Two sets of ATOMIC calculations are shown–one that includes higher-order multipole radiative transitions (i.e., those beyond the E1 transition, such as electric quadrupole [E2] and magnetic dipole [M1] radiative transitions) in the calculation, and one in which only E1 transitions are considered. We found that the inclusion of higher-order multipole transitions makes a significant difference to the spectrum, and in particular transfers population to the lowest (relativistic) configuration in each of the ion stages present at these conditions, so that fewer strong lines are observed when the multipole transitions are included. This was unexpected for two reasons. Usually, higher-order multipole transitions are expected to be important only at low ($\leq 10^{10}$ cm^{-3}) electron densities, and their inclusion typically results in more emission lines, not fewer. This study shows that the emission from W plasma at magnetic fusion conditions explores some very interesting atomic physics. The workshop comparisons showed that several bulk plasma properties (such as the average ionization of the plasma) were in reasonable agreement among the contributors, but that fine-structure calculations were required to fully characterize the emission spectra. More detailed studies of W plasma at these conditions are ongoing with ATOMIC.

In conclusion, we have reported on detailed atomic kinetics calculations that were contributed to the NLTE-7 Code Comparison Workshop. Such calculations are instructive as a verification and validation of our kinetics-modeling capabilities, and also allow us to form collaborations within this community. We aim to contribute to NLTE workshops that are planned in the future.

[1] Fontes, C.J. et al., *High Energ Density Phys* **5,** 15 (2009).
[2] Magee, N.H. et al., *Proc 14th Topical Conf Atom Process Plasma,* **168** (2004).
[3] Bastiani-Ceccotti, S. et al., *High Energ Density Phys* **3,** 20 (2007).
[4] Ralchenko, Yu. et al., *Proc 16th Int Conf Atom Process Plasma* **1161,** 242 (2009).

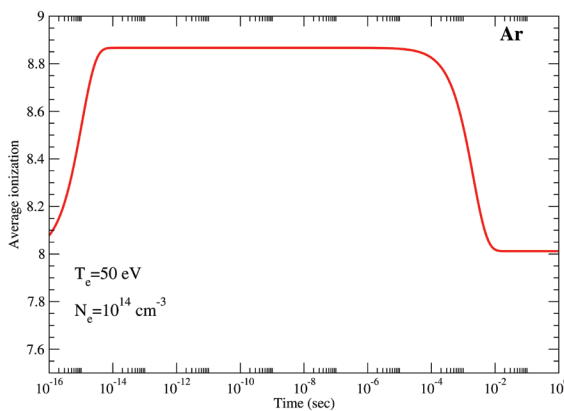# LANL Contributions to the NLTE-7 Code Comparison Workshop–
Argon Test Case

James P. Colgan, T-1; Christopher J. Fontes, XCP-5;
Gregory S. J. Armstrong, T-1; Honglin Zhang, XCP-5;
David P. Kilcrease, Joseph Abdallah, Jr., T-1

We report on the LANL contributions to the recent NLTE-7 Code Comparison Workshop that was held in Vienna, Austria in December 2011. The purpose of the workshop is to provide a detailed comparison among kinetics codes that simulate the properties of non-local-thermodynamic-equilibrium (NLTE) plasmas. Comparisons are made for a range of plasma properties such as ion populations and radiative losses. The LANL contributions were made using the LANL suite of atomic physics codes and the ATOMIC kinetics code, and included calculations for all of the test problems. In general, the LANL contributions compared very well to the other submissions, which came from various kinetics codes from institutions worldwide.

Fig. 1. ATOMIC calculations of the average ionization of Ar plasma as a function of time, at an electron temperature of 50 eV and an electron density of 10^{14} cm^{-3}. The initial condition of the plasma is that all population is in the $1s^1$ $2s^2$ $2p^6$ $3p^1$ levels of Ne-like Ar (corresponding to an average ionization of 8.0 at time=0).

The NLTE Code Comparison Workshops [1] were first organized more than 10 years ago in an effort to provide critical comparisons between atomic kinetics codes that model populations of non-local-thermodynamic-equilibrium (NLTE) plasmas and their radiative properties under a range of plasma conditions. Such comparisons are particularly useful since there are few (if any) experiments with which to compare such kinetics calculations in large portions of the range of plasma conditions of interest. The Code Comparison Workshops are held every few years, and typically attract from 10 to 15 different code contributions from institutions worldwide. Usually a small number of test cases are chosen, with the intention of making a very detailed comparison among the different calculations at the workshop within a highly focused interactive forum. A dedicated website (hosted by Yu. Ralchenko at the National Institute of Standards and Technology [NIST]) was constructed to facilitate detailed code comparisons. These workshops are very useful to the atomic physics effort at LANL, since they provide important verification and validation of our atomic physics and kinetics modeling codes, which is the principal set of codes used to generate local-thermodynamic-equilibrium (LTE) and NLTE emissivities and opacities.

The seventh workshop, which was held in Vienna, Austria in December 2011, compared four test cases, each of which offered some insight into plasma properties at various conditions. The LANL contributions to the workshop were made using the LANL suite of atomic physics codes and the ATOMIC kinetics code [2]. The calculations were made in a close collaboration of T-1 personnel (Gregory Armstrong, James Colgan, David Kilcrease, and Joe Abdallah) and XCP-5 staff members (Chris Fontes and Honglin Zhang). Much of the work performed for this workshop is expected to result in journal publications. In this report, we summarize the argon test case, with particular emphasis on the LANL calculations that were made. We also briefly describe the conclusions reached at the workshop.

The argon (Ar) test case for the NLTE-7 workshop was divided into two parts. Steady-state calculations were made for a range of electron temperatures (10 to 100 eV) and at two electron densities (10^{10} and 10^{20} cm^{-3}) to perform a detailed analysis of the role of dielectronic recombination and excitation-autoionization in the population kinetics. This case used a well-defined list of electronic configurations to facilitate the best comparison possible among the different submissions. The ATOMIC calculations for this test case, made using the configuration-average approximation, found that for a low electron density (10^{10} cm^{-3}) it was necessary to include configurations with principal quantum number up to n=18 to obtain fully converged autoionization and recombination rates. For the higher electron density case, comparisons were complicated by the need to include continuum lowering in the kinetics calculations–that is, lowering of the ionization thresholds of the atomic species due to plasma density effects.

The second Ar test case was a model time-dependent calculation inspired by the selective K-shell photoexcitation that can be accomplished using existing X-ray Free Electron Lasers (FEL). The
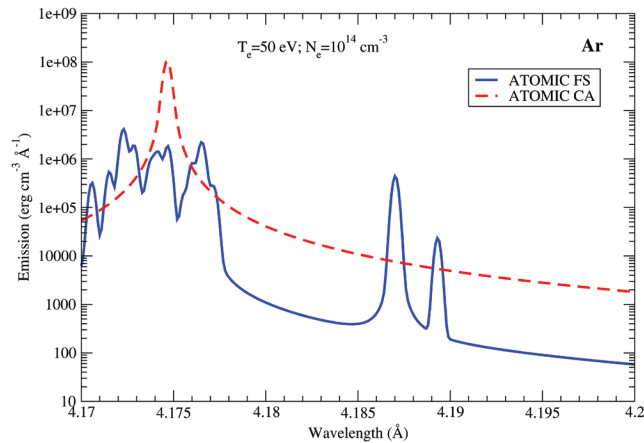
Fig. 2. ATOMIC calculations of the time-integrated emission spectrum of Ar plasma in an X-ray wavelength range, for an electron temperature of 50 eV and an electron density of 10^{14} cm^{-3}. The time-dependent kinetics calculations were made for the same initial conditions as in Fig. 1. We present configuration-average (red line) and fine-structure (blue line) calculations of the emission spectrum.

test case called for three calculations, at an electron temperature of 50 eV, and at three different electron densities (10^{14}, 10^{18}, and 10^{20} cm^{-3}). All cases were propagated in time with the initial condition that all population is assumed to be in the levels of the $1s^1 2s^2 2p^6 3p^1$ ($\equiv 1s^{-1} 3p^1$) configuration of Ne-like Ar. The ATOMIC contribution for this case again consisted of configuration-average and fine-structure calculations. Figure 1 shows the ATOMIC fine-structure calculations of the average ionization of the Ar plasma, as a function of time, for the 10^{14} cm^{-3} electron density case. This case exhibits some unusual features–in particular it was found that the plasma takes a relatively long time (0.01 s) to reach a steady state.

We found that the plasma, starting with all populations in the Ne-like stage, rapidly ionizes so that by a time of 10^{-14} s the average ionization rises to around 8.8. This behavior is due to the rapid autoionization of the $1s^{-1} 3p^1$ levels. However, about 25% of the destruction of these levels is also due to radiative decay from the 2p to 1s sub-shells, which gives rise to the X-ray spectra in the 4.17 to 4.2 Å wavelength range (see the time-integrated spectrum in Fig. 2). By a time of around 10^{-14} s, the $1s^{-1} 3p^1$ levels have completely depopulated. At this point the majority of the population is in the $1s^2 2s^2 2p^4 3p^1$ levels of fluorine (F)-like Ar. These levels decay via a two-step process (involving two E1 transitions) to the ground level of F-like Ar, and there is also some recombination into the Ne-like ion stage. The soft X-ray spectral region of 40-50 Å contains many spectral lines due to the 3l-2l' transitions in F-like Ar. At a time of 10^{-9} s (for the 10^{14} density case), most of the population is in the two ground levels of F-like Ar ($2p^5\,^2P_{3/2,1/2}$) and in the ground level of Ne-like Ar ($2p^6\,^1S_0$). These levels retain population for a sizeable time, until recombination eventually forces all population back to the ground level of Ne-like Ar, at which time the plasma is in a steady state with an

average ionization of 8.02. This latter value was confirmed by an explicit steady-state calculation.

Figure 2 shows the time-integrated emission spectrum of the Ar plasma, again for the 10^{14} cm^{-3} case. The ATOMIC fine-structure calculation is compared to the ATOMIC configuration-average calculation. This comparison shows clearly that the more detailed fine-structure calculation results in many more features in the emission spectrum in this X-ray wavelength range. As mentioned previously, the features in this region are due to transitions from the 2p to 1s sub-shells in Ne-like Ar, that is, transitions between levels arising from the $1s^1 2s^2 2p^6 3p^1$ configuration to levels arising from the $1s^2 2s^2 2p^5 3p^1$ configuration. The ATOMIC calculations were found to be among the most detailed of those presented at the workshop, and also underscored the importance of accurate atomic structure in constructing large-scale atomic kinetics calculations.

In conclusion, we have reported on detailed atomic kinetics calculations that were contributed to the NLTE-7 Code Comparison Workshop. Such calculations are instructive as a verification and validation of our kinetics-modeling capabilities, and also allow us to form collaborations within this community. We aim to contribute to NLTE workshops that are planned in the future.

[1] Fontes, C.J. et al., *High Energ Density Phys* **5,** 15 (2009).
[2] Magee, N.H. et al., *Proc 14th Topical Conf Atom Process Plasma,* 168 (2004).

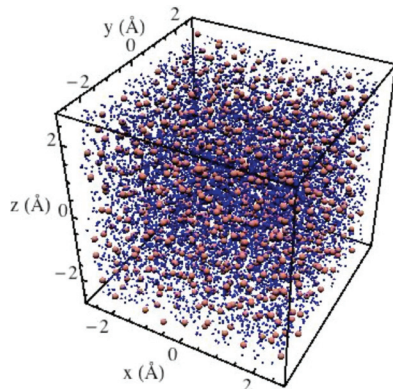# Large-scale Molecular Dynamics Simulations of Dense Plasmas

Christopher A. Fichtl, Matthew T. Calef,
Edward D. Dendy, Michael S. Murillo, CCS-2;
James N. Glosli, David F. Richards,
Frank R. Graziani, LLNL

Fusion reactivities are typically computed using a binary collision picture in which two charged particles collide to form fusion-reaction products. Here, we have examined how that basic result is modified due to nearby particles that are present in dense-plasma environments. We have utilized the massively parallel molecular dynamics code ddcMD, which we have recently ported to the Cielo supercomputer, to study these dense-plasma effects in the context of fast ignition. We illustrate the effects of the dense-plasma environment by comparing our data against predictions given by a simple theory for low-density plasmas.

Dense plasma physics plays a central role in a wide variety of astrophysical settings and laboratory experiments. The microphysical properties of such systems can be accurately computed using molecular dynamics (MD) simulations, which correctly capture the N-body collisions inherent to strongly coupled (highly collisional) plasmas by self-consistently following time evolution of individual plasma particles in a 6D phase space (see Fig. 1). We have recently ported the Gordon Bell prize-winning, massively parallel MD code ddcMD [1] to Cielo, which is currently the sixth-largest supercomputer in the world. The ddcMD code was originally developed at LLNL and has been recently adapted to dense-plasma physics by a multi-institutional team including the authors of this review. It utilizes a highly optimized particle-particle-particle-mesh (PPPM) algorithm in order to efficiently handle the Coulombic interaction between the N particles in the simulation domain [2].

Our initial investigations have involved the exploration of structural properties of many-body Coulomb systems, and their implication for fusion burn at the National Ignition Facility (NIF). In particular, we wish to understand the role of charged-particle impurities on thermonuclear burn characteristics of the dense thermonuclear fuel. Under typical NIF burn conditions the fuel is composed of a mixture of charged particles which create, and are affected by, local charge density fluctuations. These fluctuations exert forces on individual plasma ions and change the likelihood of the rare event in which a pair of atoms are close enough to one another to achieve fusion. A fundamental understanding of this process is central to understanding the subtleties inherent to the fusion process.

Two-particle events are modified in the presence of the surrounding dense plasma. To observe these effects, we examine the pair correlation function ($g(r)$) of the plasma system. This function is formally defined as the probability of finding a particle within in a shell of width dr at a distance r from a reference particle; a simple cartoon of this is shown in Fig. 2. The simplest measure of $g(r)$ for an ensemble of plasma particles is known as the "bare" result, $g_0(r)$, which considers an idealized system of only two particles interacting with each other through an unscreened potential. This model serves as a baseline to which we can compare the pair correlation data obtained directly from ddcMD simulations. The MD data contains the pair correlation effects of all the N-body interactions within the system, thus we can obtain the enhancement due to the screening inherent to the dense plasma system by defining the enhancement factor as the limit as r goes to zero of $E(r) = g(r)/g_0(r)$ [3]. In the context of fusion plasmas, enhancements greater than unity as r approaches zero indicate an increased likelihood of a fusion event occurring because particles are more likely to be close together than in an idealized system.

To illustrate the importance of the enhancement in dense plasma systems, we have run an MD simulation of a plasma with initial parameters relevant to fast ignition processes [4]. In fast ignition, a spherical deuterium-tritium (D-T) fuel pellet is first compressed and then ignited by separate laser pulses. A thin gold cone is often inserted into the D-T pellet to allow the ignition beam to more efficiently deposit its energy into the center of the pellet by "guiding" the beam and keeping this channel clear of the D-T plasma that is formed by the heating beam. As such, we have modeled a homogenous D-T mix at 100 eV and

Fig. 1. A typical MD simulation box containing particles of two species. Individual atoms of the higher-density, lower-charged species are shown in blue and the lower-density, higher-charged species are shown in red. Data is extracted from the MD simulation by averaging the particle positions and velocity distributions as the simulation advances in time.
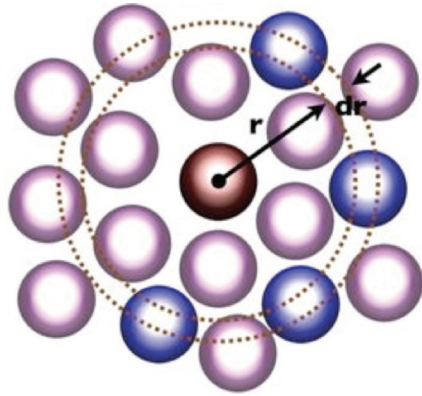
Fig. 2. Cartoon illustrating the pair correlation function g(r). Also known as the radial distribution function, g(r) is a measure of the probability of finding an atom in an annulus of width dr at a distance r from a particular atom, taken to be at the origin.

Fig. 3. Pair correlation function g(r) for D and T atoms for an 11,000-atom simulation of typical fast ignition parameters. To better illustrate the physical meaning of the pair correlation function, we have here plotted g(r,θ).
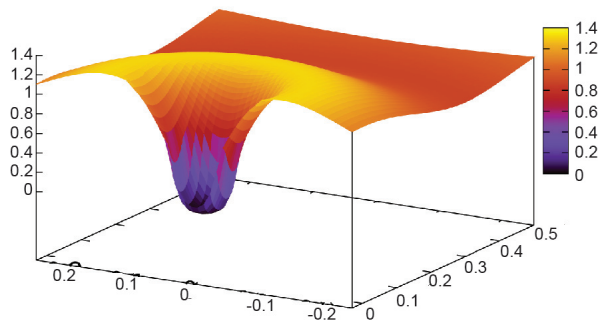
300 g/cc with a 5% (by number density) dopant of gold atoms at the same temperature. Under these conditions, we expect very strong coupling to occur between the gold atoms, as well as enhancements to g(r) at small r for the hydrogenic species. While we have successfully run plasma simulations on Cielo with 5 million atoms on ~4500 processors, and on LLNL's Blue Gene with 2.8 billion atoms on ~278,000 processors [1] in this example we have used 11,000 atoms on 10 processors–this corresponds to a cube of homogenous plasma with an edge length of 5.25 Å. The self-consistent evolution of the particles was followed for 450,000 time steps (350 fs) to reach an equilibrium configuration and collect sufficient particle statistics to construct the pair correlation function for the hydrogenic species shown in Fig. 3. The structure following the sharp incline in the radial distribution can be attributed to the presence of the gold atoms in the system.

Figure 4 shows the enhancement factor E(r) obtained from this simulation. The asymptotic enhancement value of ~6 as r approaches zero indicates that the likelihood of a fusion event occurring is increased by this amount relative to the result predicted by the binary interaction theory. This increase can be attributed directly to the influence of the dense-plasma environment. We note that the data is noisy at small r–this is to be expected because particles passing quite close to each other are rare events. We have recently developed a technique that allows us to easily sample these rare events, and as such, can measure the enhancement factor at extremely small r-values quite accurately [5]. This simple example clearly demonstrates both the influence of dense plasma effects on fusion-plasma environments and the extended (and new to LANL) capability of the MD code ddcMD.
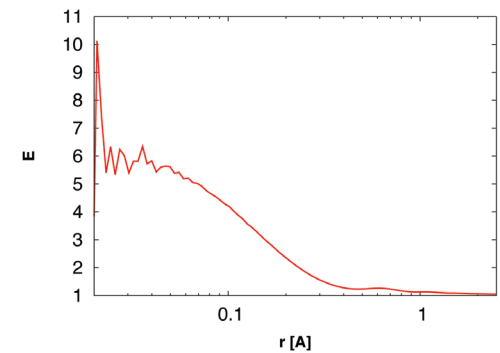
Fig. 4. The enhancement factor E(r) obtained using typical fast ignition parameters. Since like-charged particles repel each other, there are few statistics at small values of r; this leads to the noise observed in E(r).

[1] Richards, D.F. et al., *Proc Conf High Perform Comput Network Storage Anal* (2009).
[2] Graziani, F.R. et al., *High Energy Density Phys* **8,** 105 (2012).
[3] Ichimaru, S., *Rev Mod Phys* **65,** 255 (1993).
[4] Tabak, M. et al., *Fusion Sci Tech* **49,** 254 (2006).
[5] Fichtl, C.A. et al., in preparation (2012).

Common Framework for Quantifying Weapon System Uncertainty and Reliability within DOE/NNSA

Aparna V. Huzurbazar, Christine Anderson-Cook,
CCS-6

The LANL Statistical Sciences Group, CCS-6, has a long-standing effort in research and applied methods to support the assessment of conventional and nuclear weapon stockpiles. This work has been funded long term by the Enhanced Surveillance Campaign (ESC). ESC is one of the Department of Energy/National Nuclear Security Administration's (DOE/NNSA) Engineering Campaigns that brings both an engineering and science-based predictive capability to aging of nuclear weapons and maintains critical capabilities and expertise. CCS-6 provides statistical and analytical support such as system modeling, age-aware models, tracking and trending data, and uncertainty quantification. ESC work helps the Stockpile Stewardship Program, Core Surveillance, and the Nuclear Weapons System Programs. The opportunities for statistical work encompass a broad range of areas including Bayesian statistics, reliability methods, experimental design, design and analysis of computer experiments, and complex computer modeling and simulation. This article describes our work with reliability and quantification of margins and uncertainties (QMU).

Nuclear weapons are designed with multiple objectives that include safety, security, and reliability. Here we limit our focus to reliability, which is defined as the probability of success of the weapon performing its intended function at the intended time given the required temperature range, shock and vibration exposures, altitude and speed of the release envelope, and over the designed lifetime of the weapon. The goal of the Department of Energy/National Nuclear Security Administration's (DOE/NNSA) weapon reliability assessment process is to provide a quantitative metric for this assessment. Ideally, the methods to estimate this metric should be as common as possible within the nuclear weapon complex. Researchers at LANL and SNL (New Mexico and California) worked together on a reliability uncertainty aggregation team funded by the Enhanced Surveillance Campaign (ESC, C8-LANL), Joint Munitions Program (LANL), and the Advanced Certification Campaign (SNL) from fiscal year 2008 to fiscal year 2010 to develop methodology, some of which is described here. The impact of this work includes applications to Department of Defense (DoD) systems and the current B61 Life Extension Program (LEP).

Fig. 1. Statistical view of margin and threshold distributions.

Quantification of Margins and Uncertainties (QMU) is primarily a technical framework for producing, combining, and communicating information about performance margins to support risk-informed decision-making for stockpile stewardship over the nuclear weapon life cycle [1]. The elements of such an analysis require the definition of performance thresholds, calculation of performance margins, and the quantification of uncertainty about these thresholds and margins. Our interest is in methodology for uncertainty quantification (UQ) in predictions of reliability and performance. At LANL, QMU was developed to facilitate analysis and communication of confidence in assessment or certification [2]. See Fig. 1 for a statistical view of margins and thresholds and their associated uncertainty.

Our study examined a system model representative of the complexity of the top-level models used by SNL and LANL to assess weapon reliability. Given constraints and the frequently prohibitive cost of collecting data for evaluating system reliability, alternative statistical analyses were developed to leverage the understanding gained from component and sub-system level data. It has been relatively straightforward to combine multi-level data to obtain a single point estimate of system reliability, subject to assumptions about how to combine those sources based on series or parallel structures for combining the components in the system. However, appropriately propagating the uncertainty associated with each reliability estimate based on limited testing and available data has been a more difficult challenge.
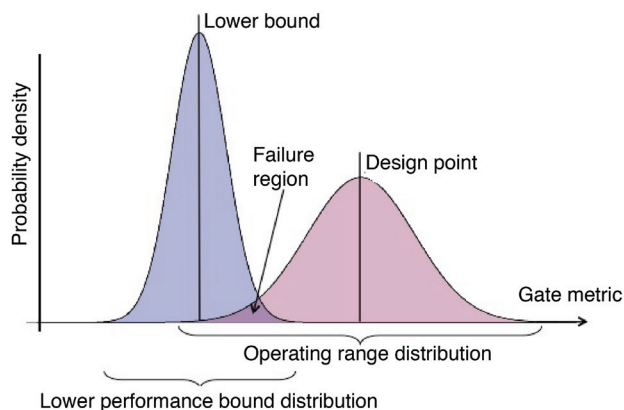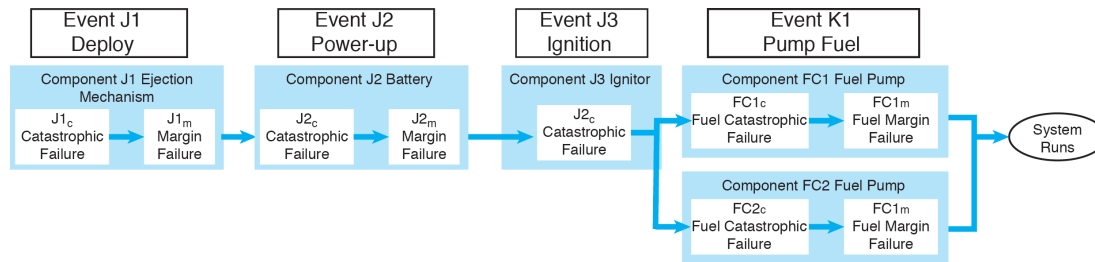
Fig. 2. System model for illustrative example in reliability block diagram format.

With a mature stockpile, limited production opportunities, and a now-extensive surveillance database, an understanding of the residual uncertainty associated with known and measurable failure modes has grown in importance. A key feature of the QMU reliability analyses is to focus not only on a point estimate of reliability, but also on the uncertainty associated with the estimate. Understanding uncertainty can and should influence subsequent decision-making. Thus the three methods presented in [3] seek to complement the NNSA point reliability and to provide a unified mechanism for assessing system-level reliability (point estimates) and reliability uncertainty (interval estimates) associated with component-level catastrophic failures and margin failures.

The authors and their collaborators, Stephen Crowder, John Lorio, James Ringland (SNL) and Alyson Wilson (Institute for Defense Analyses), developed and described different approaches for capturing reliability estimation uncertainty [3]. Three approaches are compared:
1) a historically based Method of Moments, which makes assumptions about the shape of the distribution characterizing the reliability estimate based each of the different data types and then uses these assumed distributions to combine them into an overall system estimate;
2) a Bayesian approach, which allows for related information to be combined with the observed data for each data type in order to obtain a posterior distribution, which can then be used to propagate the uncertainty between the different levels of the system model; and
3) a Bootstrap approach, which approximates the estimated component level reliabilities with an empirical distribution, and which can be

sampled to combine and give an overall reliability estimate at the system level.

The different assumptions of each method, as well as their relative strengths and weaknesses are compared. With an illustrative case study and simulated data, the three approaches are implemented and compared to show common features and differences. Figure 2 shows the modeled system structure. More details about the Bayesian approach and how it can be broadly applied to more complex scenarios where multiple data sources and multiple failure modes occur on individual components are given in [4].

[1] Diegert, K. et al., *"Toward a More Rigorous Application of Margins and Uncertainties within the Nuclear Weapons Life Cycle; A Sandia Perspective,"* SAND2007-6219 (2007).
[2] Sharp, D. and Wood-Schultz, M., *Los Alamos Science* **28,** 47 (2003).
[3] Anderson-Cook, C. et al., *Reliab Eng Syst Saf* **96,** 1063 (2011).
[4] Wilson, A. et al., *Reliab Eng Syst Saf* **96,** 1076 (2011).

Research and Development in the Jayenne Implicit Monte Carlo Project

Todd J. Urbatsch, Allan B. Wollaber,
Kelly G. Thompson, Jeffery D. Densmore,
Gabriel M. Rockefeller, CCS-2; Timothy M. Kelley,
CCS-7

The Jayenne Project, which began in 1997, provides simulation capability for thermal radiative transfer in the X-ray regime for high-energy-density physics applications such as supernova explosions, inertial confinement fusion, and radiation flow experiments at facilities such as SNL's Z-Pinch, the Omega Facility, and the National Ignition Facility [1]. The Jayenne Project uses the Fleck and Cummings [2] Implicit Monte Carlo (IMC) method to simulate the transport of thermal radiation that is non-linearly and tightly coupled to hot material. The Jayenne Project's software is multi-dimensional, runs on Adaptive Mesh Refinement (AMR) meshes, has different parallel schemes, and provides continuously improving production-level simulation while serving as a vehicle for methods research. Some of the recent research items include more accurate sub-cell representation of emission locations, adaptive implicitness, efficiency improvements (both methods- and architecture-based), and variance reduction.

In Implicit Monte Carlo (IMC), particle emission sites need to be chosen within a spatial cell. A piecewise constant representation of the emission energy in each cell produces waves that move too fast. The Jayenne Project software currently utilizes a linear discontinuous tilt scheme based on face averages, which has worked sufficiently well but can produce inaccuracies when cells are too thick and when smaller time steps increase the fraction of volume emission particles. Longer-term research to guide the tilt development has involved analytic moment analysis for continuous transport problems on spatial grids [3]. Recent mid-term research has looked at a sub-cell wavefront representation, which has evolved to a two-segment linear representation with a variable inflection point (VIP) to represent the emission function more accurately everywhere in the problem, not just the wavefront [4]. Figure 1 shows the VIP scheme in 1D, and Fig. 2 shows the improvement in the wavespeed for a spatially under-resolved Marshak Wave.

IMC's results have associated noise, so variance reduction research is an important and recurring activity. One variance reduction technique this year was in the momentum and Eddington tensor tallies [5], which was stimulated by the Jayenne Project's use as the high-order solver in LANL's high-order/low-order research effort [6]. The Jayenne Project

Fig. 1. Schematics of the two-segment, variable-inflection-point representation of the emission distribution.

represents curvilinear geometries with Cartesian geometries, such as RZ geometry with a 3D flat-top wedge; the momentum is tallied in Cartesian space and then ensemble-combined to get the radial momentum. By eliminating unnecessary sampling and recasting the estimators as embedded curvilinear estimators, we drastically reduced the variance in each of these tallies. Another variance reduction effort involved material (ion) sources in conjunction with an existing variance-/work-reduction capability that refrains from sampling particles in cells whose temperature is below a cutoff. In the IMC method, material sources are sourced partially into the radiation as part of the time-implicitness. Normally, the user can apply a cutoff so that particles are not wasted in large, cold, unimportant regions. If there is any sort of source in any spatial cell, though, we currently disable the cutoff to help conserve energy. We found that, in certain cases, doing this actually detrimentally increased the variance. We have implemented a new capability such that, in any cell below the cutoff that will not go unstable, we bypass the IMC time-implicit treatment and deposit the material source into the material. As the cell heats up, we transition back to the normal IMC material/radiation splits. This new approach reduces variance without sacrificing the stability of the IMC method.

IMC can be slow, and there are several existing algorithms that speed up the IMC. We have been developing a Discrete Diffusion Monte Carlo (DDMC) method that couples with the IMC and more cheaply moves a particle according to a diffusion approximation in regimes where the
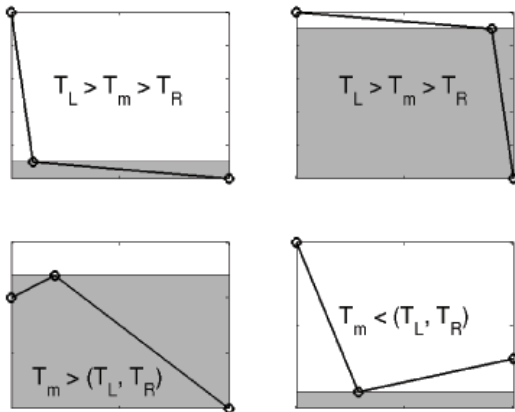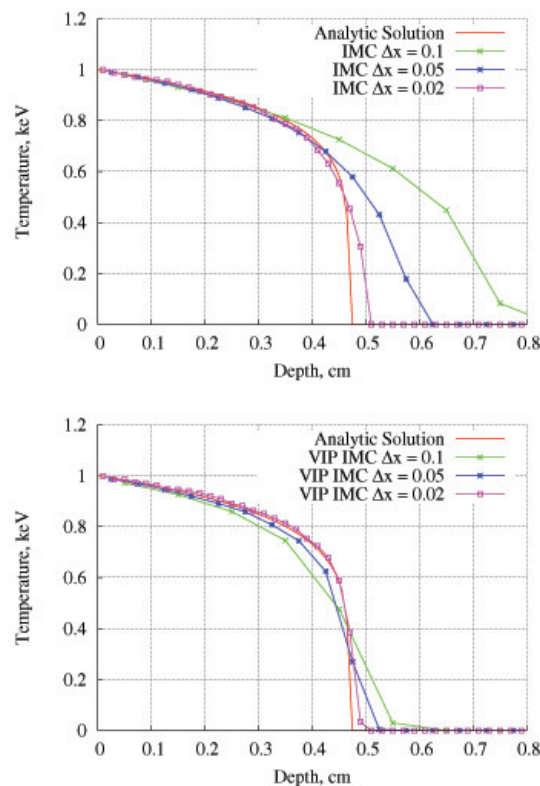
Fig. 2. Marshak wave profiles at t=10 sh using $\Delta t=0.001$ sh for the current (top) and variable inflection point (bottom) tilt schemes with varying spatial grids.

diffusion approximation is valid. The DDMC method has been shown to be up to a few orders of magnitude faster than the current state of the art. This year, the DDMC method was extended to a (frequency) multigroup model where particles whose frequencies are below a "thick" cutoff frequency are transported according to DDMC [7].

The Jayenne Project was one of the first software projects running on Roadrunner, and its adaptation focused on RZ geometry. In the summer of 2011, the 3D capability was adapted to Roadrunner with the goal of getting the most out of Roadrunner's last year of production. Looking beyond Roadrunner, we have begun exploring IMC on graphics processing units (GPU). A first project was to investigate accelerating our source calculation, which involves several loops over the spatial cells. The source calculation is a serial operation in spatial-domain replication parallelism and, in some problems, can quickly dominate runtime with increasing numbers of processes. The most important outcome from this exercise was that a new software component, GPU Device, was added to our underlying component library. This component wraps both CUDA (NVIDIA's parallel computing architecture) runtime and driver application programming interfaces (API) to provide an object-oriented view of GPU hardware and GPU kernels. The modularity of this component allows unit testing of low-level GPU capabilities isolated from other portions of the Jayenne Project codes. The new GPU component was used in creating a GPU-optimized version of the Jayenne Project's Source_Builder algorithm. During this exploration of heterogeneous programming techniques, streaming SIMD extension (SSE) vectorization by hand and OpenMP threads under message-passing interface (MPI) were also explored. For the source builder prototype, the SSE vectorization showed the best performance improvement, but the cost of moving the data to and from the GPU

still overwhelmed the speedup. All of these approaches, each of which requires different coding of the same algorithms, fit nicely into our CMake build system, requiring only subdirectories of the accelerator-specific coding. Moving beyond simple loops, we are using mini-applications to investigate the best design for rewriting our actual particle transport on GPUs.

[1] Urbatsch, T.J. and T.M. Evans, "Milagro Version 2: An IMC Code for Thermal Radiative Transfer: Capabilities, Development, and Usage," LANL Technical Report, LA-14195-MS, available at http://www.osti.gov/bridge (2005).
[2] Fleck, J.A. Jr. and J.D. Cummings, *J Comput Phys* **8,** 313 (1971).
[3] Densmore, J.D., "Spatial Moments of Continuous Transport Problems Computed on Grids," *Trans. Theor y Stat. Phys.* accepted (2011).
[4] Urbatsch, T.J. et al., "Linear sub-cell re presentations of emission in particle-based thermal radiation transport," LANL Technical Memorandum CCS-2:2011-57(U), LA-UR-11-06881 (2011).
[5] Wollaber, A., "Implementing Variance Reduction in IMC Momentum Tallies," LANL Technical Memorandum CCS-2:11-10(U) (2011).
[6] Park, H. et al., "A Consistent, Moment-Based Multiscale Solution Approach for Thermal Radiative Transfer Problems," *Trans. Theory Stat. Phys.* submitted (2011).
[7] Densmore, J.D. et al., "A Hybrid Transport-diffusion Monte Carlo Method for Frequency-dependent Radiative-Transfer Simulations," *J Comput Phys*, to be submitted (2012).

# Improved Methodology for Estimation, Uncertainty, and Validation for Licensing of Evolving Nuclear Reactors

Cetin Unal, D-DO; Brian J. Williams, CCS-6;
Christopher J. Stull, INST-OFF; David M. Higdon,
CCS-6

A methodology that can potentially be used to address new challenges in the design and licensing of evolving nuclear technology has been developed [1]. The methodology is generic and can be used as a certification/licensing framework in other technology areas including defense, energy and environment, and basic and applied research and science projects. The main components of the proposed methodology are verification, validation, calibration, and uncertainty quantification–steps similar to the components of the traditional US Nuclear Regulatory Commission (NRC) licensing approach with the exception of the calibration step. An enhanced calibration concept is introduced here, and is accomplished through data assimilation. The new methodology suggests a formalism to quantify an adequate level of validation (predictive maturity [2]) with respect to existing data so that additional experimental testing can be minimized, reducing costs by demonstrating that this testing will not enhance the quality of the predictive tools.
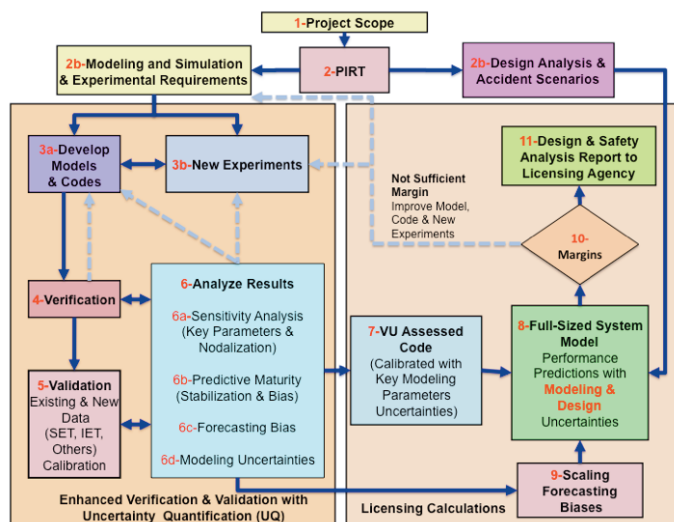
Application of the advanced validation methodology of Fig. 1–involving verification, validation, calibration, and uncertainty quantification–to the nuclear fuel performance codes FRAPCON and LIFEIV was performed for the Nuclear Energy Advanced Modeling and Simulation (NEAMS) program. FRAPCON is used to predict oxide fuel behavior in light water reactors (LWR). LIFEIV was developed in the 1980s to predict oxide fuel behavior in fast reactors. We utilize a screening methodology to narrow down the selected parameters for sensitivity and calibration analyses. We deployed the screening methods to both codes and discussed results. The number of modeling parameters selected was 61 for FRAPCON and 69 for LIFEIV. Screening resulted in only 24 parameters of importance for FRAPCON, while the LIFEIV analysis reduced important modeling parameters to 34.

LIFEIV sensitivity studies indicated that the fuel thermal conductivity and gas release models were most influential in terms of explaining overall output variability after calibration to available pin data, and are therefore targets for additional calibration to further constrain their parameters. For example, Fig. 2 shows the change in average fission gas release induced by calibrated marginal variation in each of 18 LIFEIV calibration parameters, along with a measure of variability in fission gas release induced by these calibrated single-parameter effects. Residual (post-calibration) uncertainties in parameters TC2 (fuel thermal conductivity) and FGPM1 (fission gas release) have the most pronounced effects on fission gas release, and are thus candidates for further uncertainty reduction.

The gap thermal conductance and crack elasticity models had a less pronounced effect on overall output variability. We combined these results with the results of the screening study to recommend a ranking of models that can be considered for further improvements. Our model ranking is as follows:

- fuel creep;
- fuel conductivity;
- fission gas transport/release;
- crack/boundary; and
- fuel gap conductivity.

We think a major review of the fuel creep model and uncertainties associated with its parameters is needed urgently. Means of calibrated parameter distributions were compared with nominal LIFEIV values,
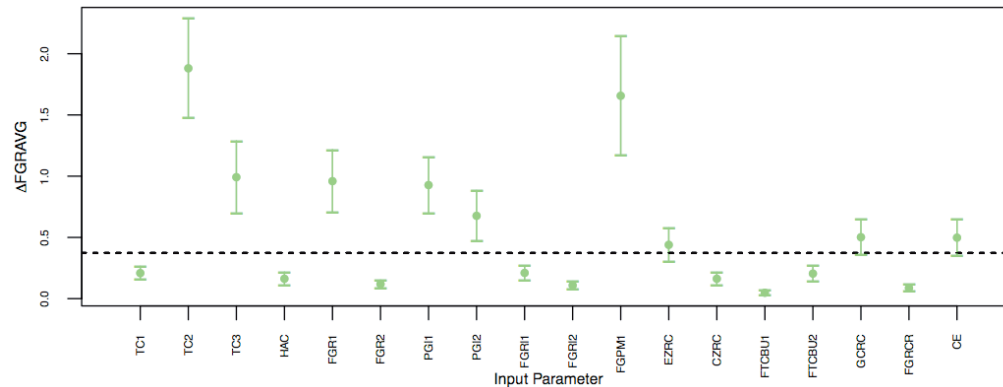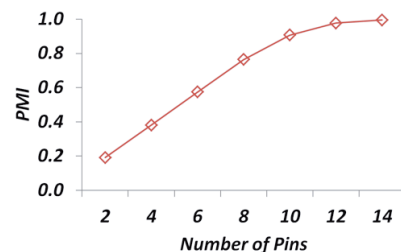

Fig. 1. Overview of the improved best estimate plus uncertainty methodology.

Fig. 2. Post-calibration sensitivity analysis of 18 LIFEIV calibration parameters (x-axis) on average fission gas release (y-axis).

Fig. 3. LIFEIV PMI as a function of the number of experiments.

indicating that changes of up to ±10% can be observed. This effort will continue with assessment of calibrated model predictions against additional independent validation pin data.

FRAPCON sensitivity results primarily agreed with results obtained from LIFEIV and indicated that the fuel thermal conduction and fission gas release models are two key modeling areas on which to concentrate for further reduction of uncertainty. The following particular parameters were found to be important contributors.

- Thermal Conduction
  - phonon term in thermal conduction model;
  - porosity correction to thermal conduction accounting for radiation effects; and
  - overall fuel thermal conductivity.
- Fission Gas Release
  - multiplier in grain boundary accumulation model;
  - saturation area density gas multiplier;
  - burn-up enhancement factor applied to diffusion constant; and
  - diffusion constant multiplier .
- Design Parameters and Indirect Modeling Parameters
  - pod surface heat flux at elevation z on the rod axis; and
  - as-fabricated fuel-cladding gap size.

The four fission gas release calibration parameters were further investigated with a larger set of data. It was found that the adjustment to activation energy was the most sensitive parameter in predicting gas release fraction if other modeling parameters (especially the fuel conduction model) were not considered.

"When is a prediction from a numerical model good enough?" All too often, the answer to this question relies on expert opinion: a qualitative and subjective answer on which to rely, given the quantitative nature of verification, validation, and uncertainty quantification. To address this concern, researchers at LANL have been investigating a metric referred to as the Predictive Maturity Index (PMI). The PMI offers a succinct and quantitative mechanism by which to track year-to-year improvements of a numerical model, while also providing a description of particular areas that need improvement. The PMI currently takes four attributes into account:

- coverage of the design space over which the models are applied;
- discrepancy between the model predictions and the experimental data;
- complexity of the model, relative to the state-of-the-art models; and
- robustness of the models to lack of knowledge on the part of the analyst(s).

An example of this framework applied to the LIFEIV calibration study is given in Fig. 3 where it is seen that the addition of pins to the suite of available experimental data results in a corresponding increase in the value of the PMI.

[1] Unal, C., et al., *Nucl Eng Des* **241,** 1813 (2011).
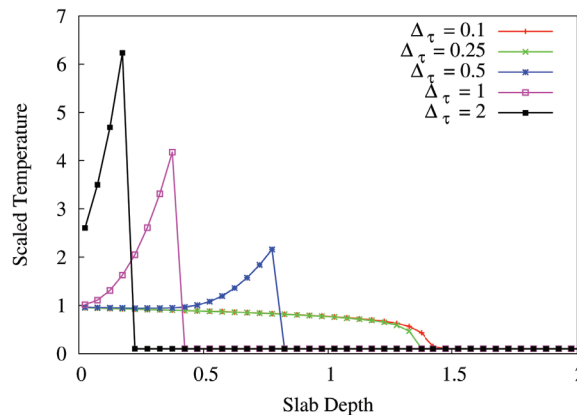[2] Hemez, F. et al., *Comput Struct* **88,** 497 (2010).

A Discrete Maximum Principle for the Implicit Monte Carlo Equations

Allan B. Wollaber, Jeffery D. Densmore, CCS-2;
Edward W. Larsen, University of Michigan;
Paul W. Talbot, Oregon State University

Over forty years ago, the Implicit Monte Carlo (IMC) equations emerged as a robust linearization of the nonlinear thermal radiative transfer equations that describe the interaction of photons with matter. However, for sufficiently large time steps, the IMC equations can produce nonphysical "temperature spikes," and the conventional remedy has been to manually reduce the time-step size until these artifacts disappear. We have derived necessary, sufficient, and approximate conditions on the time-step and temperature-grid sizes to ensure that the IMC temperature solutions satisfy a discrete maximum principle–that is, in the absence of inhomogeneous sources, the internal material temperatures remain between specified boundary conditions. This work is now being used as the basis of a dynamic time-step controller for the IMC equations.

The Implicit Monte Carlo (IMC) equations are an unconditionally stable time-discretization and linearization of the thermal radiative transfer equations that are amenable to a stochastic, particle-based (Monte Carlo) solution algorithm [1]. However, the IMC equations can admit nonphysical temperature solutions if the time step is chosen to be sufficiently large [2]. Specifically, for problems that contain no inhomogeneous sources, it has been shown that the thermal radiative transfer equations obey a maximum principle–that is, the interior temperatures are bounded above and below by the boundary condition temperatures at all times [3]. By contrast, if a sufficiently large time step is chosen, the IMC temperature solution can be made to exceed the boundary condition temperatures in a heating problem. Fig. 1 provides several examples of these IMC maximum principle violations. It depicts the spatial temperature profile of an initially cold material that is suddenly subjected to a hot, isotropic source of radiation on the left boundary at a dimensionless time of $\tau=8$ using several different choices of time-step sizes. As the time-step size is increased, the maximum temperature of this Marshak wave increasingly overshoots the unit left boundary condition and retards the wavespeed.

Twenty-five years ago, Larsen and Mercier [2] provided a sufficient condition on the time-step size of the spatially continuous IMC equations to ensure that their solution does not violate the maximum principle. However, their sufficient condition was considered too conservative for practical use, as simulations showed that time-step values that were orders of magnitude larger than their recommendation could be employed without the appearance of nonphysical temperature solutions.

Recently, we have developed new time-step restrictions to prevent nonphysical overheating that explicitly consider the spatial-grid size of the temperature discretization [4]. If the restriction on the time- and space-dependent grid parameters is satisfied, then we say that the IMC solution satisfies a discrete maximum principle. Our main approach is to approximately solve the IMC radiative-transfer equations, determine an estimate of the maximal radiation energy deposited in a mesh cell, and demand that the resulting temperature update not exceed the boundary temperature. This demand directly results in an approximate time-step recommendation such that the IMC equations do not violate the discrete maximum principle. Because our approach is approximate, we also developed rigorous necessary and sufficient conditions on the maximal time-step size. Employment of our approximate technique on sample problems thus far has provided highly accurate predictions of the grid-dependent, maximal time-step size for the IMC equations to admit physical solutions.

We tested our approach on a 1D, nonlinear, Marshak wave problem–an initially cold slab of material that is suddenly subjected to a hot, isotropic, temperature source on one boundary–using a wide range of spatial and temporal grid parameters and both frequency-dependent and grey (frequency-integrated) radiation descriptions in the Milagro

Fig. 1. Temperature profiles at $\tau=8$ for a Marshak wave problem in which the time step is varied using $\Delta_\tau=0.1$, 0.25, 0.5, 1, and 2 mean free times for emission.
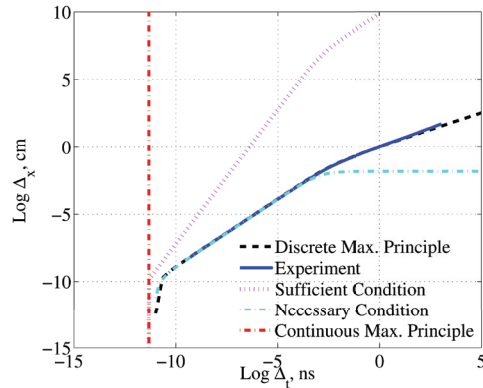
Fig. 2. First incidences of theoretically predicted (dashed) and experimental (solid) maximum-principle violations for a grey Marshak wave problem. The Δ_T from the continuous maximum principle is the vertical line at the left.
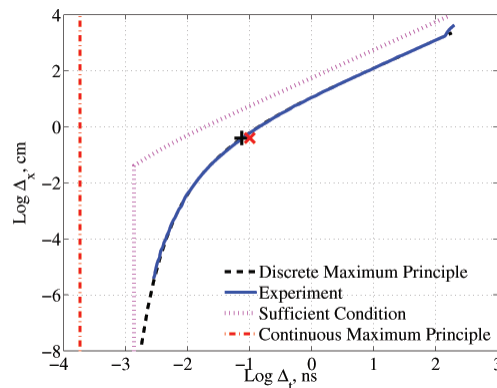

Fig. 3. First incidences of theoretically predicted (dashed) and experimental (solid) maximum-principle violations for a frequency-dependent Marshak wave problem.

IMC code [5]. The grey results are depicted in Fig. 2 on a log-log scale in which the abscissa measures the time-step size and the ordinate the spatial grid size. The solid blue curve indicates interpolated numerical results of IMC calculations that segregate physical solutions (above and to the left of this line, corresponding to small time steps and large spatial cells) and nonphysical solutions (below and right of this curve, corresponding to large time steps and small spatial cells). The black, dashed curve represents the critical time-step sizes predicted by our approximate technique. These exhibit excellent agreement. Additionally, they are bounded above and below by the rigorous sufficient and necessary conditions that we developed (dashed magenta and dot-dashed cyan lines, respectively). Finally, we note that the vertical red line corresponds with the grid-free recommendation of Larsen and Mercier [2], which can be many orders of magnitude smaller than the observed value depending on the chosen time-step size. Figure 3 depicts the results of the same experiment when frequency-dependent opacities are incorporated, which necessitated the development of more involved approximations [4]. Again, our approximate prediction (the black, dashed curve) is in excellent agreement with interpolated numerical results from many IMC simulations (the solid blue curve). The necessary condition is not shown, as it was always satisfied for this problem. Figure 3 also depicts two data points from Larsen and Mercier's earlier work, the black "plus" sign, corresponding to an IMC calculation without maximum principle violations, and the red "x," corresponding to a nonphysical solution. Because the left plus sign is about three orders of magnitude larger than the predicted time-step limit, Larsen and Mercier remarked that their time-step limit was too conservative.

However, Fig. 3 shows that their choice of the spatial grid size had a substantial influence on that assessment.

Because the memory requirements of the numerical solution of the IMC equations are significant, it has been computationally prohibitive to adaptively restart an IMC calculation from the previous cycle using a smaller time step whenever maximum principle violations are encountered. Using our theory as its foundation, research is now underway to construct an inexpensive, grid-dependent, dynamic time-step controller for the IMC equations that should preclude nonphysical temperature spikes in the IMC temperature solution. Paul W. Talbot of Oregon State University has undertaken the 3D extension and implementation of this work into Jayenne Project software as part of his Master's project [6].

[1] Fleck, J.A. Jr., and J.D. Cummings, *J Comp Phys* **8,** 313 (1971).
[2] Larsen, E.W. and B. Mercier, *J Comp Phys* **71,** 50 (1987).
[3] Andreev, E. et. al., *Zh.vych. Mat. i mat. Fiz* **23,** 52 (1983).
[4] Wollaber, A.B. et. al., *Nucl Sci Eng,* to appear (2012).
[5] Urbatsch, T.J. and T.M. Evans, LANL Technical Report LA-14195-MS (2005).
[6] Talbot, P.W. et. al., *Trans Am Nucl Soc* **106** (2012)
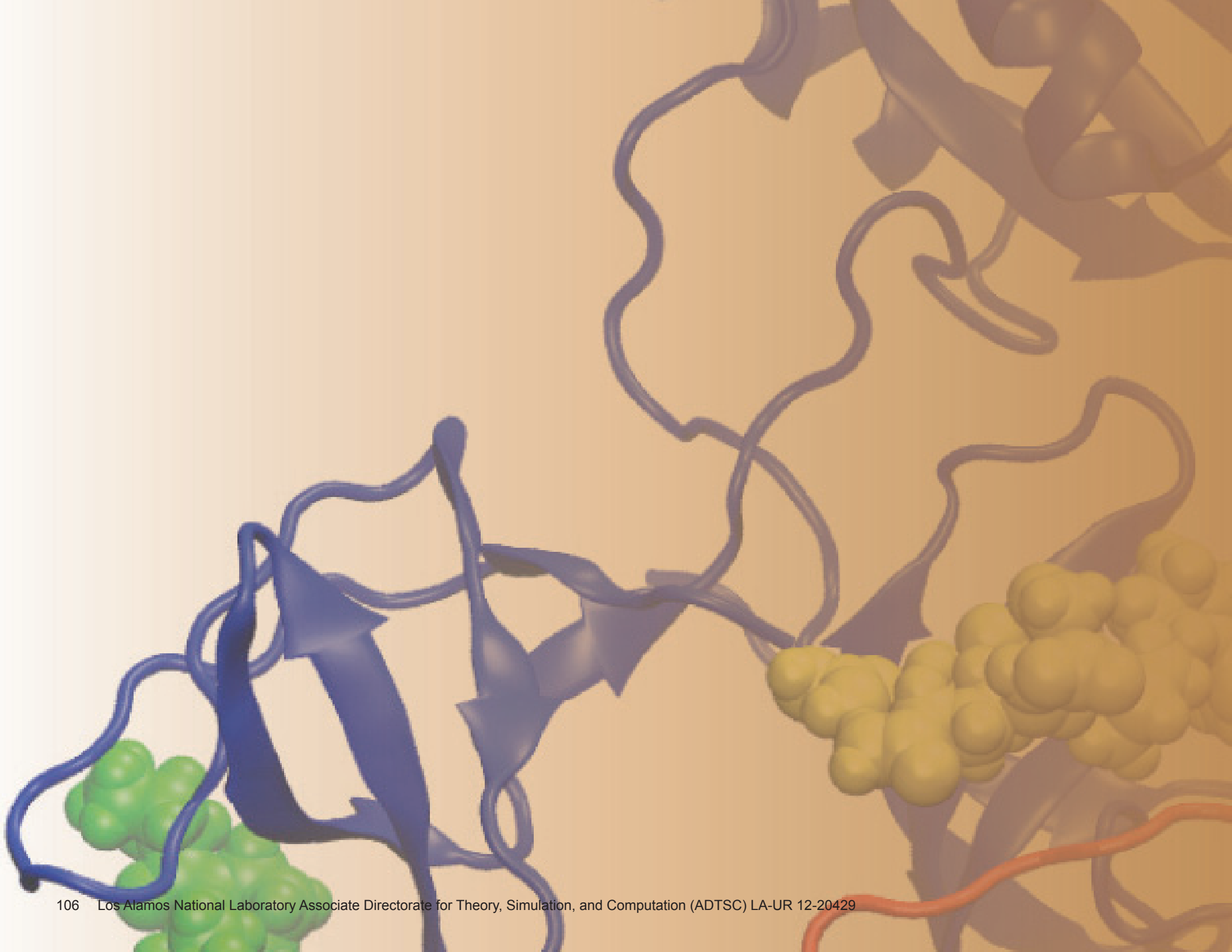
Chemistry and Biology

LANL has a long history of outstanding work in chemistry and biology. This work is dependent upon the cross collaborations of experimentalists here at LANL as well as academia and other national laboratories.

Articles in this section cover research in cell signaling, DNA dynamics, biofuels, HIV transmission, disorder in proteins, and noncoding RNA.

Quantifying Intramolecular Binding in Multivalent Interactions:
A Structure-based Synergistic Study on the Grb2-Sos1 Complex

Anurag Sethi, Byron Goldstein, S. Gnanakaran, T-6

The formation of multivalent complexes by combining relatively weak promiscuous interactions is a strategy that is often used to increase the affinity and specificity of biomolecular complex formation among signaling proteins. However, there is often no clear structural information regarding multivalent interactions between signaling protein partners and their effect on the stoichiometry and the distribution of these protein complexes. We developed a hybrid method that combines molecular dynamics simulations and polymer models to estimate the intramolecular concentration of a motif near the vacant binding site in the same complex during multivalent binding. We developed and applied this method to the study of complex formation between the adaptor protein growth receptor binding protein-2 (Grb2) and son of sevenless-1 (Sos1). Using this method, we were able to estimate the stoichiometry and distribution of complexes formed under physiological conditions.
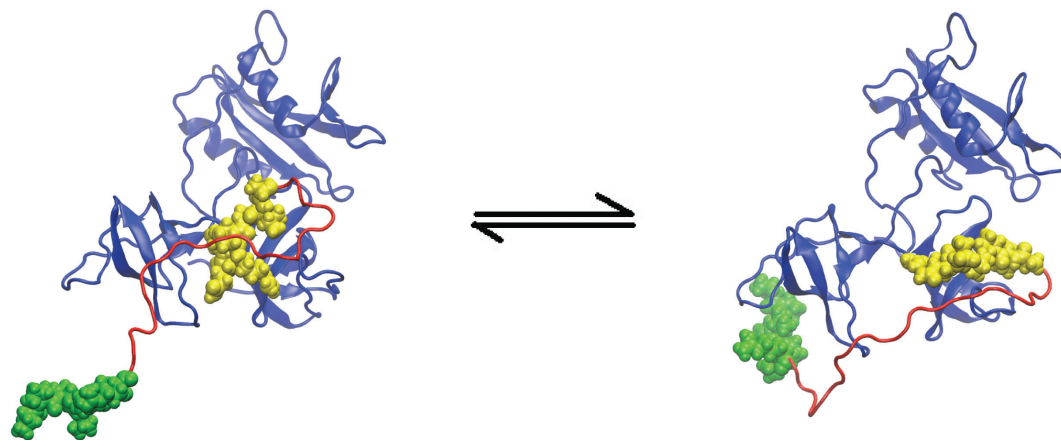
Many signaling proteins use multivalent interactions in which relatively weak interactions are combined to increase the affinity and specificity of complex formation. Often, in such interactions, one of the biomolecules (protein A) involved consists of a disordered region that contains multiple ligands (or motifs) that can each bind to the binding sites of a protein partner (protein B). Often the protein, in our case growth receptor bound protein-2 (Grb2), is bivalent. When one of the binding sites of protein B is bound to a motif on protein B, the two molecules are tethered to each other. This results in an increase of the local concentration of the free motifs on protein B near the free site(s) on protein A. The related intramolecular binding constants and effective concentrations are very difficult to determine experimentally. In order to estimate the intramolecular binding constant(s) theoretically, polymer models are often used to model the disordered regions in the biomolecule. However, the flexibility of the protein that contains the binding site(s) is ignored. We developed a hybrid approach combining molecular dynamics simulations and polymer modeling that accounts for the flexibility in both molecules to estimate the intramolecular binding constants [1].

We applied this method to the study of the multivalent binding between the adaptor protein Grb2 and son of sevenless-1 (Sos1) and provided a rationale for the stoichiometry of the complexes that were observed experimentally. Grb2 contains two SH3 domains (small protein domain of about 60 amino acid residues) that interact with multiple polyproline motifs separated by flexible unstructured regions on Sos1. Grb2 mediates the recruitment of Sos1 from the cytosol to the plasma membrane where it activates Ras (a family of related proteins involved in transmitting signals within cells) by inducing the exchange of guanosine diphosphate (GDP) for guanosine triphosphate (GTP). First, using a combination of evolutionary information and binding-energy calculations, we predict an additional polyproline motif in Sos1 that binds to the SH3 domains of Grb2. This gives rise to a total of five polyproline motifs in Sos1 that are capable of binding to the two SH3 domains of Grb2. Then, combining molecular dynamics simulations and polymer models, we estimate the enhancement in local concentration of a polyproline motif on Sos1 near an unbound SH3 domain of Grb2 when

Fig. 1. The intramolecular binding of a polyproline motif (green) on Sos1 to the SH3 domain of Grb2 (blue) after another polyproline motif (yellow) in the same Sos1 molecule is bound. The two polyproline motifs in Sos1 are connected by a linker (red).
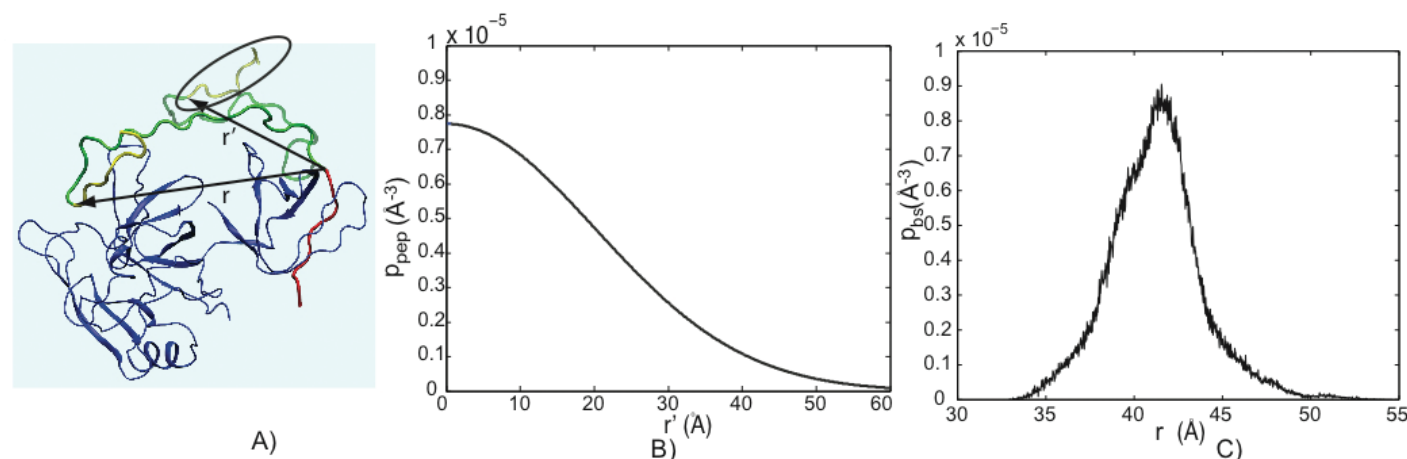
Fig. 2. (a) An SH3 domain on Grb2 (blue) is exposed to a greater concentration of a polyproline motif (yellow) on Sos1 (circled) after one of the motifs (red) in Sos1 is bound to the other SH3 domain in Grb2. The linker region connecting the two motifs is shown in green. In the hybrid model, we use a (b) worm-like chain model to calculate the probability density of the distance between the two ends of the linker, while a (c) molecular dynamics simulation is used to measure the probability density of the distance between the two binding sites in Grb2.

its other SH3 domain is bound to a different polyproline motif on Sos1. We show that the local concentration of the Sos1 motifs that a Grb2 SH3 domain experiences is approximately 1000 times greater than the cellular concentration of Sos1. Finally, we calculate the intramolecular equilibrium constants for the crosslinking of Grb2 on Sos1 and use thermodynamic modeling to calculate the stoichiometry. With these equilibrium constants, we were able to predict the distribution of complexes that form at physiological concentrations. We believe this is the first systematic analysis that combines sequence, structure, and thermodynamic analyses to determine the stoichiometry of the complexes that are dominant in the cellular environment.

This is a general method that can be used to model many multivalent complexes observed in biology, such as in signaling molecules, antibody-antigen binding, host-pathogen interactions, and DNA-transcription factor interactions. In addition to estimating the intramolecular binding constants, we also gained a molecular understanding of the nature of complexes that are formed in these systems. These principles could be used for rational design of drugs or vaccines that would optimally bind in a multivalent fashion based upon such models.

[1] Sethi, A. et al., *PLoS Comput Biol* **7**, e1002192 (2011).

DNA Conformation Dynamics and Human Diseases

Boian S. Alexandrov, T-1; Alan R. Bishop, PADSTE; Luba Kalaydjieva, Western Australian Institute for Medical Research; I. S. Kohane, Childrens Hospital, Harvard Medical School; Kim Ø. Rasmussen, T-5; Anny Usheva, Beth Israel Deaconess Medical Center; V. I. Valtchinov, Harvard Medical School

The inherent thermal motion that a DNA molecule is subjected to in a living cell makes it possible for the two strands of the molecule to locally and spontaneously open and close. This dynamic effect is referred to as "DNA breathing" because it induces local transient separation of complementary base pairs. Numerous prior studies that combine modeling and experiments have firmly established a connection between this localized breathing dynamic of DNA and functional characteristics of mammalian transcriptional promoters. Here we highlight two recent studies that utilized the relationship between the breathing dynamics of DNA and biological function to elucidate some of the mechanisms of pathogenesis and progression in cognitive deficit schizophrenia, Friedreich's ataxia, Huntington's disease, and Fragile X-syndrome.

Fig. 1. DNA breathing dynamics and gel-shift experiments. The examined DNA sequence is displayed at the top with the SNPs in red and the transcription start site in green. (a) Gel-shift assay with nuclear extracts. Unlabeled oligonucleotides used in the self- and cross-competition assays are shown on top, with triangles indicating increasing concentrations (20–100-fold molar excess) of the competitor. Lanes: 1, no nuclear extract; 2, no competitor; 3 and 4, self-competitor; 5 and 6, cross-competitor. Bands: NS, nonspecific; A and B, specific, with decreasing intensity in the self-competition experiment. Preferential binding of the protein complex to the C-G sequence is suggested by the inability of the G-A oligonucleotide to out-compete the labeled C-G probe. (b) Langevin dynamics simulations of DNA breathing. Strong breathing activity can be seen for a sequence containing the C-G sequence (left), in contrast to the poor breathing activity of the G-A sequence (right) [5].

In living cells DNA molecules are subjected to inherent thermal motion, which makes it possible for each DNA base pair, or for a number of consecutive base pairs, to spontaneously open and re-close. This effect is referred to as "DNA breathing" because it induces local transient openings of the double helix. Numerous studies have shown that this local conformational dynamic of DNA is important for genomic functions such as transcription and replication. Previous research studies that combine modeling and experiments [1–4] have firmly established a correlation between the local breathing dynamics of DNA (also known as bubbles) and functional characteristics of mammalian transcriptional promoters. The results suggest that the breathing dynamics of DNA
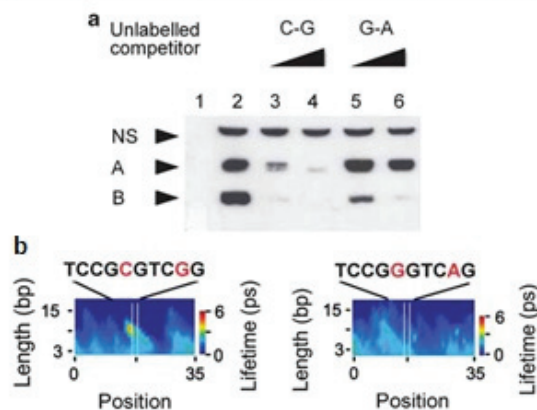
can more generally be correlated with regulatory genomic regions, and as such can be important for biological function. It is, however, uncertain whether breathing dynamics might be a determining factor in pathogenesis or in a disease state. Therefore, it is important to examine whether genomic mutations that compromise local bubble patterns may enhance the risk of a particular disease. Such genomic mutations are not necessarily located in regulatory genomic regions. Their disruption of biological function (e.g., suppressed transcription) may also be caused by long-range changes of the physicochemical or structural properties of the double helix. Importantly, various changes of the nucleotide sequence can lead to a similar impairment in conformational breathing dynamics. Thus, seemingly inconspicuous variations in DNA sequence may result in disease onset or progression. Here, we highlight two recent studies that utilized the relationship between the breathing dynamics of DNA and biological function to elucidate some of the mechanisms of pathogenesis and progression in cognitive deficit schizophrenia [5], Friedreich's ataxia, Huntington's disease, and Fragile X-syndrome [6].

Cognitive deficit schizophrenia. In a recent study published in *Molecular Psychiatry* [5], we studied the effect of single-base-pair variations (single nucleotide polymorphisms [SNP]), which have been implicated in cognitive deficit in schizophrenia, on DNA dynamics and gene expression. Specifically, we report experimental evidence of gene expression combined with modeling of DNA breathing dynamics that points to two adjacent promoter SNPs as the functional variants associated with schizophrenia. Our modeling revealed significant
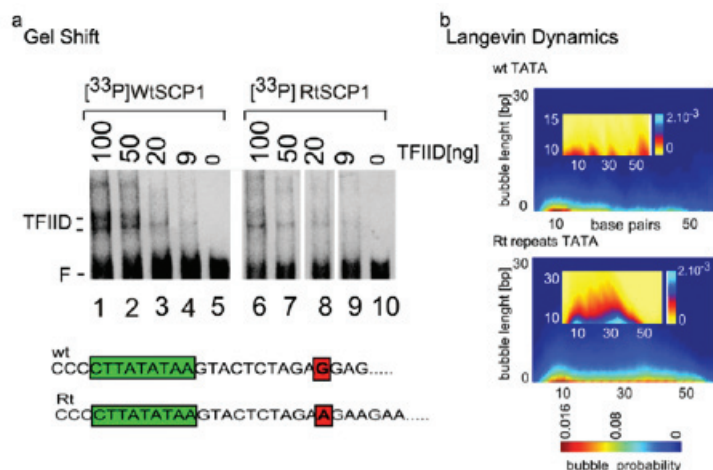
Fig. 2. Effect of TRSs on protein DNA binding and local DNA dynamics. DNA wild type (wt) and mutated (Rt) sequences displayed at the bottom. Protein binds to the segment in green. The Rt has been modified to contain 15 (GAA.TTC) repeats downstream of the TATA box as indicated. a) Band shift titration reactions received double-stranded oligonucleotides containing the shown sequences. The reactions in lanes 1–4 and 6–9 received different amounts of purified protein as indicated at the top of the lanes. The reactions in lanes 5 and 10 did not receive protein. The positions of the free DNA (F), and transcription factor (TFIID) are indicated on the left. b) Langevin dynamics simulations of the DNA breathing dynamics in Wt and Rt sequences [6].

sequence effects of the SNPs on the local breathing dynamics of DNA (Fig. 1). In particular, strong breathing activity was predicted for a sequence containing the C–G sequence (Fig. 1, left) in contrast to the poor breathing potential of the G–A sequence (Fig. 1, right), which causes a preferential binding of the protein complex to the C–G sequence and changes the gene expression. Gel shift and transcription experiments confirmed the conclusions from our simulations (Fig. 1) [5].

Trinucleotide repeat disorders. Friedreich's ataxia, Huntington's disease, and Fragile X-syndrome are examples of genetic diseases associated with the expansion of trinucleotide repeat sequences (TRS). We have found that in TRSs, the local breathing spectrum is altered by the increased symmetry that results from the repeats. It is commonly accepted that the TRS expansion is caused by the formation of non-B-DNA structures that disrupt normal cellular processes. In addition, it is believed that the initiation of non-B-DNA structures can be caused or enhanced by transient DNA openings. Hence, we simulated opening profiles of TRSs with various lengths and GC-contents. Our simulations suggested that large transient DNA bubbles can be formed not only in A/T-rich sequences (e.g., (GAA.TTC)n, Friedreich's ataxia), but also in sequences with a relatively high G/C-content, such as (CTG. CAG)n in Huntington's and (CGG.CCG)n in Fragile X-syndrome. Our simulation data implicate the transient DNA openings as a common factor in the trinucleotide repeat disorders. Further, we found that TRS sequences tend to open more cooperatively as the length of the repeats increase. This collective behavior leads to abrupt bubble creation that significantly enhances the presence of intermediate DNA bubble states,

as compared with sequences with a random nucleotide composition or with much shorter repeat tracts. In summary, we have demonstrated that TRS expansion in the disease-related (longer) sequences leads to enhanced coherent DNA openings – that is, enhanced local strand separations – when compared with the "healthy" sequences with a low number of repeats. Finally, we provided an experimental example of how such enhanced TRS bubble spectra can interfere with protein-DNA interactions, and thus with biological function. We demonstrated that the flankings composed of repeats have a profound effect on the spectrum of the TATA-box DNA-breathing dynamic, which could explain the lost TFIID-TATA binding (Fig. 2). This novel correlation between the transient bubble spectrum and repeats expansions in the individual genomes suggests local DNA dynamics as an "epigenetic" determinant that plays an important role in the TRS genetic diseases.

[1] Bishop, A.R. et al., *Springer Series in Materials Science* **148,** Kakeshita, T., et al. Eds., Springer, New York (2012).
[2] Alexandrov, B.S. et al., *PLoS Comput Biol* **5,** e1000313; PMID: 19282962 (2009).
[3] Choi, C. et al., *Biophys J* **95,** 597; PMID: 18390611 (2008).
[4] Alexandrov, B.S. et al., *Nucleic Acids Res* **37,** 2405; PMID: 19264801 (2009).
[5] Jablensky, A. et al., *Mol Psychiatr,* doi: 10.1038/mp.2011.129; PMID: 21968932 (2011).
[6] Alexandrov, B.S. et al., *PLoS One* **6,** e19800; PMID: 21625483 (2011).

Unlocking Biomass Energy: A Computational Study on Cellulose Recalcitrance

Giovanni Bellesia, T-6/CNLS;
Ramakrishnan Parthasarathi, T-6; Antonio Redondo,
T-DO; Shishir Chundawat, Bruce Dale, Great Lakes
Bioenergy Center; S. Gnanakaran, T-6

Cellulose, an abundant component of lignocellulosic biomass, is self-assembled in plant cell walls as crystalline nanofibers. It can be hydrolyzed to monomeric glucose that can eventually be converted to biofuels (e.g., alcohols, alkanes) via microbial fermentation or chemical catalysis. The degradation of cellulose polymers to sugar monomers is currently an inefficient and expensive process. We used multiresolution computer simulations to investigate the molecular origin of cellulose recalcitrance to degradation. In particular, we focused on two cellulose crystalline forms that display different degrees of recalcitrance to degradation: cellulose I-beta (the natural crystalline form of cellulose) and the less recalcitrant cellulose III(I) obtained from cellulose I-beta by treatment with liquid ammonia.

Fig 1. Cellulose I-beta and III(I) crystal structures.

Cellulose I$_\beta$
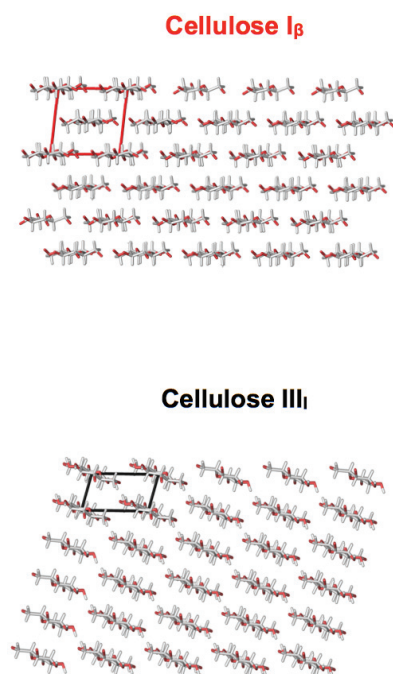
Cellulose III$_I$

One way to greatly improve the enzyme degradation process of crystalline cellulose fibrils to glucose is to convert the naturally occurring crystalline form of cellulose (cellulose I-beta) to a different crystalline form named cellulose III(I) (see Fig. 1). The conversion process from crystalline cellulose I-beta to cellulose III relies on a chemical treatment based on anhydrous liquid ammonia. Recent experiments show that the enzymatic degradation rate is two to five times greater in cellulose III(I) than in cellulose I-beta [1].

We performed a set of extensive fully-atomistic molecular dynamics (MD) simulations to 1) gain a detailed understanding of the main structural and hydration differences between cellulose I-beta and cellulose III-I, and 2) relate these differences to the different enzyme degradation rates in these two cellulosic substrates [1]. Our MD simulations revealed substantial differences in the structural stability, surface properties, and hydration between cellulose I-beta and cellulose III(I) fibrils. These results helped greatly in explaining the molecular foundations of crystalline cellulose stability. In particular, the comparative computational analysis of both the crystalline core and the solvent-exposed surface of the cellulose I-beta and III(I) fibrils identified key structural and molecular features in cellulose III(I) that can explain its lower recalcitrance and its enhanced hydrolysis rates [1].

Additional MD simulations were instrumental in providing the first detailed, mechanistic model for the interaction of liquid ammonia with native cellulose and for the consequent structural crossover to cellulose III(I) [2]. We observed the rapid formation of an extended hydrogen-bond (HB) network between liquid ammonia and the cellulose molecules on the surface of the fibril. This newly formed HB network induces a major shift in the conformation of the cellulose chains within the crystalline fibers and drives a structural crossover from cellulose I-beta to a new crystalline structure compatible with both the ammonia-cellulose I complex and cellulose III(I). A pictorial view of the structural crossover is given in Fig. 2. Our simulations reveal that significant changes in both the structural and hydration properties of the cellulose fibril happen prior to ammonia percolation and suggest rational ways to improve existing cellulose chemical pretreatments.

In a subsequent study, we combined classical MD simulations and quantum calculations to study the details of the interplay between hydrogen bonding and stacking interaction, as well as their role in the stabilization of crystalline cellulose I-beta and cellulose III fibrils [3]. The results demonstrate that individual HB interactions are stronger in cellulose I-beta than in cellulose III(I). However, the total HB contribution to stabilization is larger in cellulose III(I) due to the highly cooperative nature of the intersheet HB network. In addition, a significant contribution to the stabilization of cellulose I-beta from cooperative stacking interactions was observed. On the one hand, these results serve as a first attempt to quantify the interplay between HB and stacking interactions in cellulose. On the other hand, the results also suggest that both HB and stacking interactions need to be taken into account when designing novel chemical pretreatments and enzymes for improved catalysis of cellulosic biomass for biofuel production.
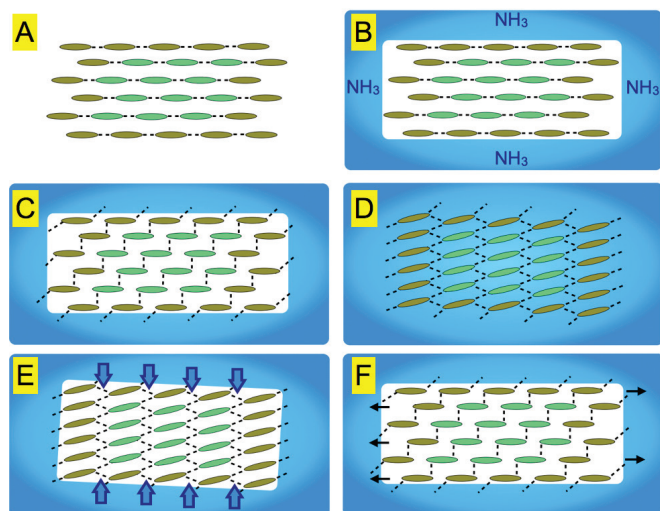
Fig. 2. Mechanistic model for the structural crossover from cellulose I-beta to cellulose III(I).

Finally, we introduced a generic coarse-grained molecular model for crystalline cellulose whose relevant degrees of freedom have been identified from the analysis of our fully atomistic MD simulations (see Fig. 3). The model provides a simple explanatory framework for structural transitions in crystalline cellulose in general, and for the transition from cellulose I-beta to cellulose III(I) in particular, where the search for optimal spatial arrangement of the cellulose chains appears to be more important than the chemical details of the cellulose crystals. In terms of catalytic cellulose deconstruction for biofuels production, the model connects the lower recalcitrance of cellulose III(I) (when compared to cellulose I-beta) to higher free energy and to overall larger structural flexibility. The model suggests that the experimental strategies with the greatest chance of succeeding as effective chemical treatments for cellulose are those based on chemical compounds that specifically interfere with the energetics of the intersheet distance (the distance between neighboring horizontal sheets in cellulose fibrils–see Fig. 1) and compounds that expand cellulose I-beta via alteration of an internal rotational degree of freedom have the most chance of succeeding as effective chemical treatments for cellulose [4].
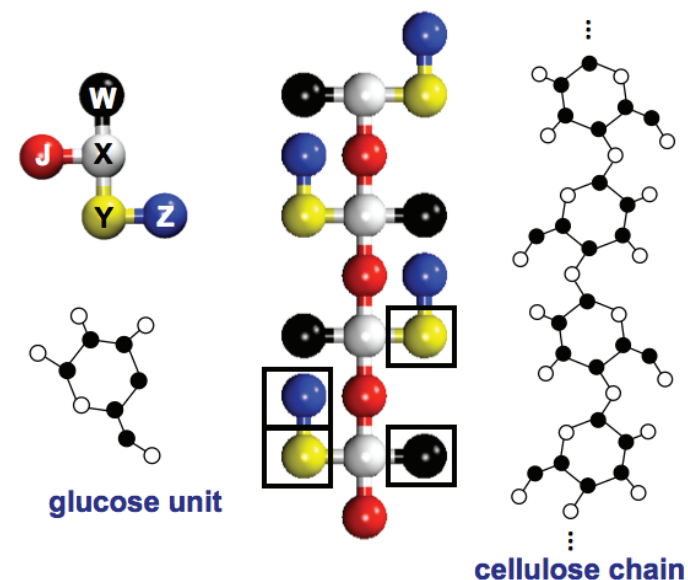

Fig. 3. Simplified geometry of our generic coarse-grained model for crystalline cellulose. Each glucose monomer (left) is represented by five beads: two for the backbone (X and J; white and red, respectively) and three for the side chains (Y, W and Z; yellow, black and blue, respectively).

[1] Chundawat S.P.S.*, Bellesia G.* et al., *J Am Chem Soc* **133,** 11163 (2011).
[2] Bellesia G. et al., *J Phys Chem B* **115,** 9782 (2011).
[3] Parthasarathi R. et al., *J Phys Chem A* **115,** 4191 (2011).
[4] Bellesia G. et al., *J Phys Chem*, submitted (2012).
*shared first authorship

Role of Donor Genital Tract HIV-1 Diversity in the Transmission Bottleneck

Peter T. Hraber, Elena E. Giorgi, T-6;
Tanmoy Bhattacharya, T-2; Bette T. Korber, T-6;
Debrah Boeras, Mackenzie Hurlston, Susan Allen,
Cynthia Derdeyn, Eric Hunter, Emory University;
Tammy Evans-Strickfaden, Clyde Hart, Centers for
Disease Control and Prevention; Joseph Mulenga,
Zambia Emory, HIV Research Project,
Lusaka, Zambia; Etienne Karita, Projet San
Francisco, Kigali, Rwanda

In the heterosexual transmission of HIV-1 infection, a genetic bottleneck is imposed on the virus quasispecies. To understand whether limited genetic diversity in the genital tract (GT) of the transmitting partner drives this bottleneck, viral sequences from blood and genital fluids of eight transmission pairs from Rwanda and Zambia were analyzed. The chronically infected transmitting partner's virus population was heterogeneous, with distinct genital subpopulations. GT populations in two of four women sampled longitudinally were stable for weeks to months. Surprisingly, the transmitted founder variant was not derived from the predominant GT subpopulations. Rather, in each case, the transmitting variant was phylogenetically distinct from the sampled locally replicating population. Although the exact distribution of the GT virus population at transmission cannot be unambiguously defined in human studies, it is unlikely that the transmission bottleneck is always driven by limited viral diversity in the donor GT or that HIV transmission is purely random.

The predominant mode of HIV-1 infection is heterosexual transmission, where a genetic bottleneck is imposed on the virus quasispecies. To understand whether limited genetic diversity in the genital tract (GT) of the transmitting partner drives this bottleneck, we analyzed viral envelope sequences from the blood and genital fluids (cervical swab or semen) of eight linked transmission pairs (both donor and recipient) from Rwanda and Zambia. The chronically infected donor's virus population was heterogeneous and predominated by distinct GT subpopulations (Fig. 1). Virus populations within the GT of two of four women sampled longitudinally exhibited stability over time intervals on the order of weeks to months. Surprisingly, the transmitted founder variant was not derived from predominant genital tract subpopulations. Rather, in each case, the transmitting variant was phylogenetically distinct from the predominant locally replicating populations in the sample. Though the exact distribution of the virus population present in the GT at the time of transmission cannot be unambiguously defined in these human studies, it is unlikely that the transmission bottleneck is driven in every case by limited viral diversity in the donor GT. Further, a quantitative test of random transmission indicates that HIV transmission is not solely a stochastic sampling from the donor GT, but more likely involves selection for some property other than abundance in the GT [1].

To address whether or not the transmitted sequence is randomly sampled from the donor virus population, we used an objective clustering criterion to relate sequences. A distance threshold D defines a cluster as any subset of sequences within distance D nucleotides from one another. That is, any two sequences occupy the same cluster if they differ at D sites or fewer. We computed pairwise distances among available GT sequences from any given donor, then assigned sequences to clusters. For any given D, and for each transmission pair i=1,…,N, we calculate the frequency f_i of donor GT sequences that do not cluster with other sequences. Let $P_D(n)$ be the probability that n donors transmit a sequence outside a cluster. Then:

$$P_D(0) = \prod_{i=1}^{N}(1-f_i)$$

$$P_D(1) = \sum_{j=1}^{N}\prod_{i \neq j}f_j(1-f_i) = P_D(0)\sum_{i=1}^{N}\frac{f_i}{1-f_i}$$

and so on. In general, the probability generating function is:

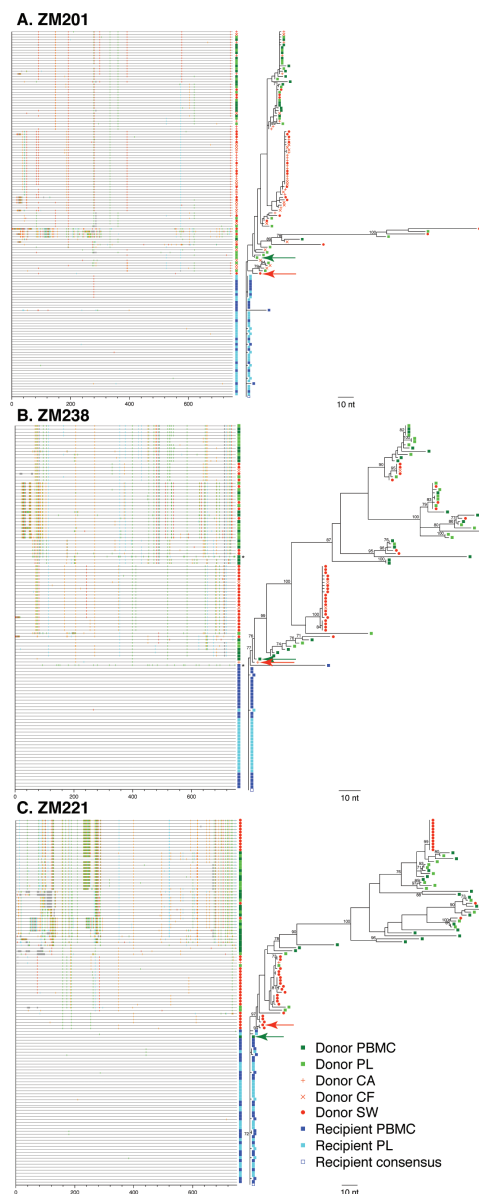$$g_D(x) = \sum_{i=1}^{N}P_D(i)x^i.$$

It follows that:

$$g_D(x) = \prod_{i=1}^{N}\left[(1-f_i) + f_i x\right].$$

We thus calculated $P_D(n)$ using the formula:

$$P_D(n) = \frac{1}{n!}\frac{d^n}{dx^n}\left[g_D(x)\right]\Big|_{x=0}$$

Fig. 1. Transmission analysis of HIV-1 with molecular phylogenetics shows distinct subpopulations in the donor female genital tract (GT). Though subpopulations of nearly identical sequences dominate female GT samples, the transmitted sequences are instead limited to distinct variants. For three of eight representative transmission pairs from Zambia (ZM) and Rwanda, aligned env V1-V4 nucleotide sequences are shown as phylogenetic trees (right) paired with highlighted polymorphism plots (left). Both GT (red symbols: CA, cell-associated; CF, cell-free; SW, cervical swab) and blood-borne virus (green squares: PBMC, peripheral blood mononucleocytes; PL plasma) from donors are shown. The recipient blood-borne virus (blue squares: PBMC, peripheral blood mononucleocytes; PL plasma) closely resembles sequences in the donor population, confirming epidemiological linkage. The most closely related donor sequences from blood (green arrows) and GT (red arrows) samples are indicated. Tick marks highlight locations of nucleotide differences (A, green; T, red; G, orange; C, cyan; gaps, grey) from the recipient consensus sequence (open blue square), the putative transmitted/founder virus in the recipient.

for $n=1,\ldots,8$ and $D=0,\ldots,9$ as above. To test for random transmission, for each D, we consider n_{obs} the observed number of donors that transmit sequences outside a cluster and compute:

$$p(D) = \sum_{n=n_{obs}}^{N} P_D(n)$$

The quantity $p(D)$ is the overall probability among eight donors that the observed number of sequences transmitted outside (not inside) a cluster is different than we would observe if transmission were to sample randomly from the donor virus population. Small values of $p(D)$ indicate that the event is unlikely to occur by chance (Fig. 2).

Software to compute the probability of non-random transmission is freely available online at ftp://ftp-t10.lanl.gov/pub/hivdb/tort/tort.tar.gz.
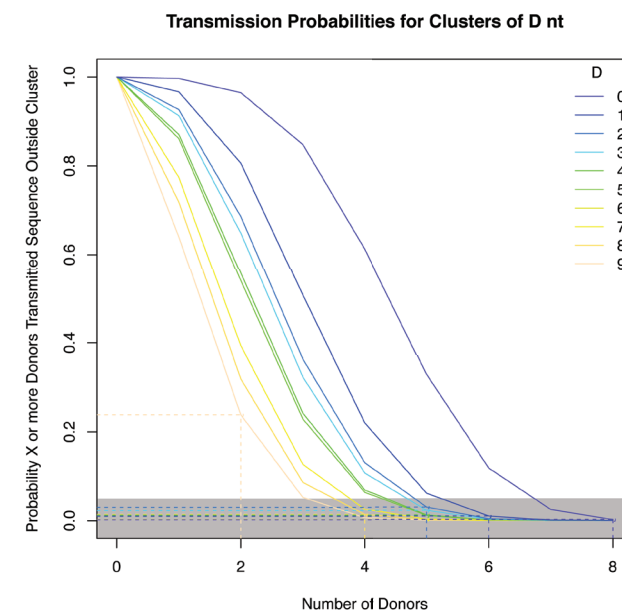

Fig. 2. Selection of transmitted/founder viruses from the donor population is not random. Subpopulations of donor GT sequences were clustered with threshold distance criterion, D, which varied from 0–9 nucleotides. For each D, the proportion of nearly identical (clustered) GT sequences was used to compute the probability that virus transmission sampled randomly from outside a cluster. Comparing these probabilities across eight transmission pairs with the number of transmission events found to occur outside a cluster (vertical dashed lines) quantifies the probability of random transmission. Small probabilities indicate non-random transmission (p<0.05, grey box).

[1] Boeras, D. et al., *Proc Natl Acad Sci USA* **108,** E1156 (2011).

Revealing Intrinsically Disordered Protein Interactions with Membrane Mimic Surfactants

Jianhui Tian, Anurag Sethi, T-6; Divina Anunciado,
Dung M. Vu, C-PCS; S. Gnanakaran, T-6

α-Synuclein (αSyn) is an intrinsically disordered protein that is implicated in Parkinson's disease. Although its physiological functions are still unknown, it has been hypothesized that it may function as a presynaptic protein due to its association with the presynaptic membrane. Recent studies have shown that αSyn can adopt different conformations that are dependent on its interaction with biomolecular aggregates such as micelles and lipid membranes. However, the exact conformation that αSyn adopts and its interaction with phospholipid environments are still unclear. In order to provide insights into αSyn structure and function we are carrying out all-atom molecular dynamics simulations of αSyn with sodium dodecyl sulfate (SDS) micelles.

The intrinsically disordered protein α-Synuclein (αSyn) is abundantly expressed in the brain [1]. It is localized at the nerve termini in close proximity to synaptic vesicles. Its native function is thought to involve vesicle maintenance and recycling, modulation of neural plasticity, endoplasmic reticulum-Golgi trafficking, and dopamine reuptake. Numerous studies connect αSyn to Parkinson's disease (PD) as it has been found to form supramolecular fibrils, which are a major component of Lewy bodies and Lewy neurites (two major hallmarks of the disease). However, it is still not clear how αSyn executes its function, and what are its toxic and key conformations for fibril formation.

In solution, monomeric αSyn has no well-defined structure but appears to be more compact than a random-coil conformation. Recently, it has also been found that native αSyn exists in cells as a helically folded tetramer. The functionality of αSyn is expected to be associated with its interaction with membranes. When interacting with lipid vesicles or membranes, αSyn typically displays two different conformations—a helix-turn-helix structure or an extended helical structure, depending on the curvature of the binding surface. On the one hand, the composition of lipid bilayers (membranes or vesicles) is known to affect the binding strength of αSyn. On the other hand, both electrostatic and hydrophobic interactions are important in the association of αSyn with

Fig. 1. The number of SDS head group atoms (OS, S, and O) (red line) and alkyl tail atoms (black line) within 0.35 nm of the protein as a function of time for αSyn started from helix-turn-helix conformation. Inset shows the same plot with time in logarithmic scale to enlarge the details of the early interactions.

lipid bilayers, and this association can lead to changes in the physical properties of bilayers.

As a lipid-membrane mimic, sodium dodecyl sulfate (SDS) has been widely used to study the role of lipid binding in modulating the conformational changes of αSyn as well as its aggregation process. A conformational interconversion between unfolded, helix-turn-helix and extended helices happens during isothermal protein-SDS titration experiments [2]. The fibrillation pathways of αSyn have been shown to differ in the presence and absence of SDS. In addition, fibrillation occurs only at low concentrations of SDS surfactant.

We are conducting extensive molecular dynamics simulations of αSyn with self-assembling and self-assembled SDS micelle systems. Our goal is to understand the details of the interactions between αSyn and SDS, and to characterize the influence of αSyn on SDS micellation.

During the simulations of the self-assembly of SDS micelles, monomeric SDS freely diffuses and interacts with the independently diffusing αSyn molecule. It binds αSyn and forms small aggregates around the whole protein. The interaction between αSyn and monomeric SDS is tracked by calculating the number of sulfate head group (OS, S, and O) atoms and alkyl chain carbon atoms within 0.35 nm of protein atoms as a function of time for all simulations (see Fig. 1).

We observed that the interaction between αSyn and SDS is not solely driven by the electrostatic interactions between the positively charged side chain of protein and the negatively charged sulfate head groups
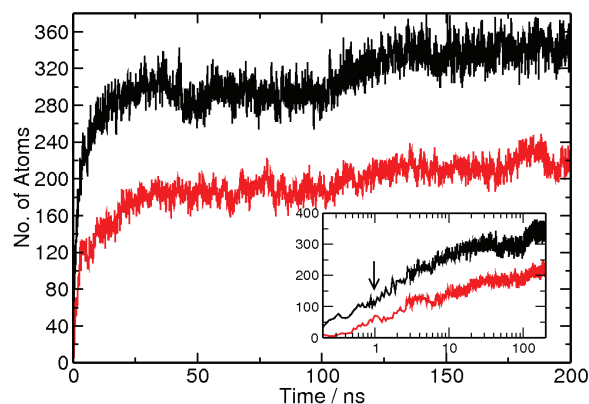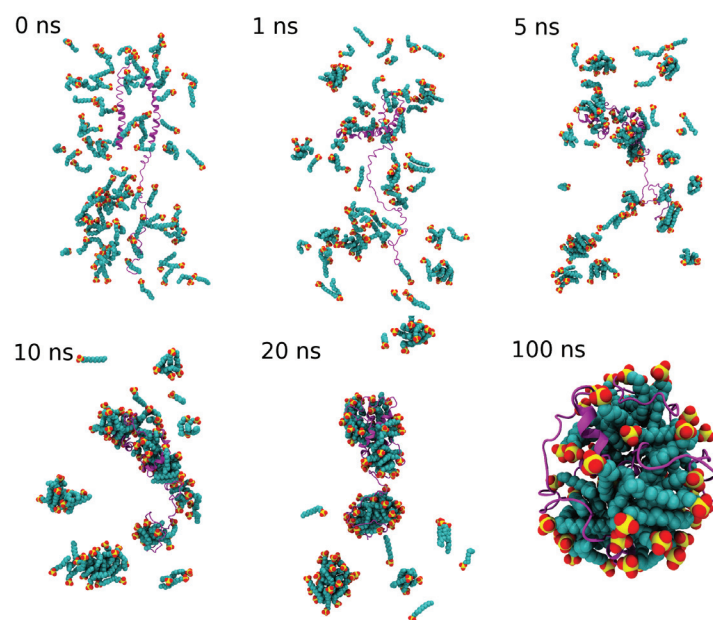
Fig. 2. Snapshots of SDS micellation at 0, 1, 5, 10, 20, and 100 ns from the 100-folded system. Only the αSyn and SDS in the final complex are shown in the snapshot for clarity. The αSyn is in magenta cartoon presentation. The SDS molecules are in space-filling representation, while the head groups are in red and yellow and the alkyl carbon tails are in cyan. For clarity, the SDS monomers, which are not part of the final complex, are not shown.

in SDS–there is also a significant contribution from the hydrophobic interaction between SDS tails and the apolar amino acids in αSyn. Our observation clearly confirms previous experimental results showing that the hydrophobic interaction plays an important role in SDS-protein interactions.

During micellation, we observe that SDS interacts with αSyn and forms a stable micelle-protein collapsed complex. As shown in Fig. 2, the micelle formation and the interaction with the αSyn occur simultaneously. We propose a three-stage process for initial micelle formation. First, monomeric SDS interacts non-specifically with αSyn and forms small aggregates along the protein chain. Second, the small aggregates grow by micelle fusion while the collapse of αSyn at the same time appears to promote the fusion process. Third, we observe the formation of a collapsed αSyn-SDS micelle complex.

We also consider non-rigorous quantification of αSyn secondary structure propensity at the residue level by calculating the average helix and beta-sheet percentage (Fig. 3). Residues 1 to 30 have a high helical structure propensity; residues 31 to 45 and 106 to 140 have no helical propensity, while residues 46 to 106 display intermediate helical propensity, with some interruptions. Residues 31 to 100 show a beta-sheet structure propensity. Three interruptions in helical propensity are observed in Fig. 3 for the N-terminus: residues 30 to 47, 66 to 69 and 83 to 87, all of which agree very well with the experimentally-measured regions with interruptions in helicity.
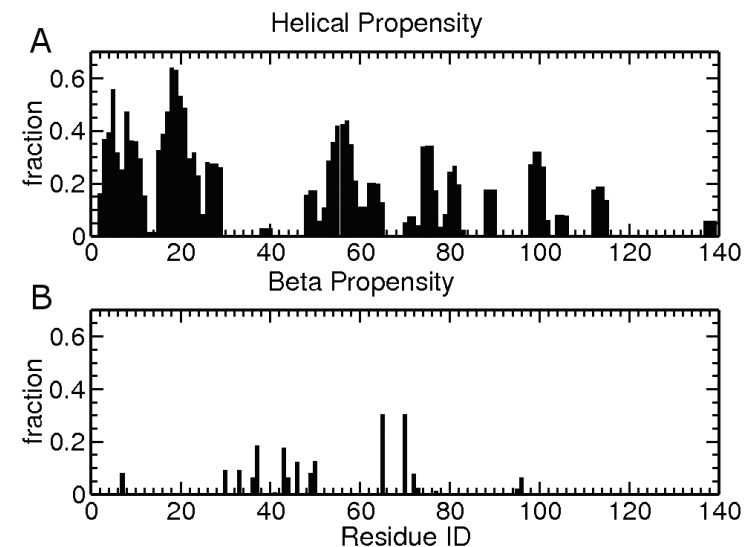

Fig. 3. Secondary-structure propensity for each of the residues in αSyn.

[1] George, J.M. et al., *Neuron* **15**, 361 (1995).
[2] Ferreon, A.C. et al., *Proc Natl Acad Sci Unit States Am* **106,** 5645 (2009).

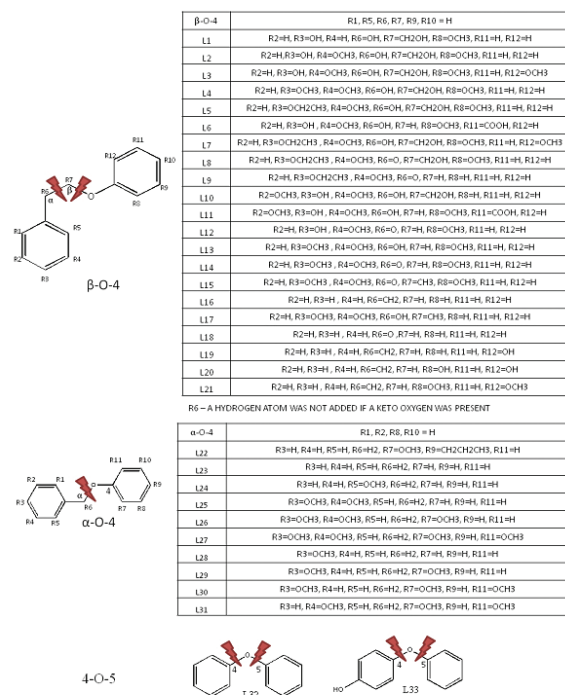# Theoretical Prediction of Bond Strength for the Diverse Set of Lignin Linkages

Ramakrishnan Parthasarathi, T-6;
Raymond A. Romero, New Mexico State University;
Antonio Redondo, T-DO; S. Gnanakaran, T-6

Lignin in plant cell walls is a potential renewable source of biofuels and chemicals. It consists of various aryl ethers, irregularly connected by a variety of linkages creating a complex structural network–thus it is difficult to identify selective bond-breaking events. In this work, we predict dissociation tendencies of a diverse set of lignin linkages using the density functional theoretical (DFT) approach. The chosen 65 lignin model compounds represent the most prevalent carbon-oxygen (ether) and carbon-carbon (C–C) bond linkages. Results from our study identify the weakest and strongest linkages connecting arene rings in different classes of lignin compounds. Also, the dissociating linkages can have different adjacent substituents, such as the methoxy group on the arene ring and hydrocarbon, methyl, and hydroxyl group substitutions on aliphatic carbon atoms. These substituents affect the ease of dissociation of lignin linkages and can be used to develop predictive models for delignification.

Fig. 1. Lignin model compounds (β-O-4, α-O-4, and 4-O-5 with ether linkages). β-O-4 contains both ether and carbon-carbon linkages.

β-O-4	R1, R5, R6, R7, R9, R10 = H
L1	R2=H, R3=OH, R4=H, R6=OH, R7=CH2OH, R8=OCH3, R11=H, R12=H
L2	R2=H,R3=OH, R4=OCH3, R6=OH, R7=CH2OH, R8=OCH3, R11=H, R12=H
L3	R2=H, R3=OH, R4=OCH3, R6=OH, R7=CH2OH, R8=OCH3, R11=H, R12=OCH3
L4	R2=H, R3=OCH3, R4=OCH3, R6=OH, R7=CH2OH, R8=OCH3, R11=H, R12=H
L5	R2=H, R3=OCH2CH3, R4=OCH3, R6=OH, R7=CH2OH, R8=OCH3, R11=H, R12=H
L6	R2=H, R3=OH, R4=OCH3, R6=OH, R7=H, R8=OCH3, R11=COOH, R12=H
L7	R2=H, R3=OCH2CH3 , R4=OCH3, R6=OH, R7=CH2OH, R8=OCH3, R11=H, R12=OCH3
L8	R2=H, R3=OCH2CH3 , R4=OCH3, R6=O, R7=CH2OH, R8=OCH3, R11=H, R12=H
L9	R2=H, R3=OH, R4=OCH3, R6=O, R7=H, R8=OCH3, R11=H, R12=H
L10	R2=OCH3, R3=OH, R4=OCH3, R6=OH, R7=CH2OH, R8=H, R11=H, R12=H
L11	R2=OCH3, R3=OH, R4=OCH3, R6=OH, R7=H, R8=OCH3, R11=COOH, R12=H
L12	R2=OCH3, R3=OH, R4=OCH3, R6=O, R7=H, R8=OCH3, R11=H, R12=H
L13	R2=H, R3=OCH3, R4=OCH3, R6=OH, R7=H, R8=OCH3, R11=H, R12=H
L14	R2=H, R3=OCH3, R4=OCH3, R6=O, R7=H, R8=OCH3, R11=H, R12=H
L15	R2=H, R3=OCH3 , R4=OCH3, R6=O, R7=H, R8=OCH3, R11=H, R12=OCH3
L16	R2=H, R3=H, R4=H, R6=CH2, R7=H, R8=H, R11=H, R12=H
L17	R2=H, R3=OCH3, R4=OCH3, R6=O, R7=CH3, R8=H, R11=H, R12=H
L18	R2=H, R3=H , R4=H, R6=O ,R7=H, R8=H, R11=H, R12=H
L19	R2=H, R3=H , R6=CH2, R7=H, R8=H, R11=H, R12=OH
L20	R2=H, R3=H , R6=CH2, R7=H, R8=OH, R11=H, R12=OH
L21	R2=H,R3=H, R4=H, R6=CH2, R7=H, R8=OCH3, R11=H, R12=OCH3

R6 – A HYDROGEN ATOM WAS NOT ADDED IF A KETO OXYGEN WAS PRESENT

α-O-4	R1, R2, R8, R10 = H
L22	R3=H, R4=H, R5=H, R6=H2, R7=OCH3, R9=CH2CH2CH3, R11=H
L23	R3=H, R4=H, R5=H, R6=H2, R7=H, R9=H, R11=H
L24	R3=H, R4=H, R5=H, R6=H2, R7=H, R9=H, R11=H
L25	R3=OCH3, R4=OCH3, R5=H, R6=H2, R7=H, R9=H, R11=H
L26	R3=OCH3, R4=OCH3, R5=H, R6=H2, R7=OCH3, R9=H, R11=H
L27	R3=OCH3, R4=OCH3, R5=H, R6=H2, R7=H, R9=H, R11=OCH3
L28	R3=OCH3, R4=H, R5=H, R6=H2, R7=H, R9=H, R11=H
L29	R3=OCH3, R4=H, R5=H, R6=H2, R7=OCH3, R9=H, R11=H
L30	R3=OCH3, R4=H, R5=H, R6=H2, R7=OCH3, R9=H, R11=OCH3
L31	R3=H, R4=OCH3, R5=H, R6=H2, R7=OCH3, R9=H, R11=OCH3

Plant cell walls consist of cellulose, hemicellulose, and lignin, all of which provide shape, elasticity, and rigidity, as well as protection, to plants. Lignin is one of the most abundant naturally occurring aromatic biopolymers and acts as the essential glue that gives plants their structural integrity and fortification. The potential of lignin is yet to be fully realized and depends on the development of selective transformations of the complicated lignin network. Lignin results from the polymerization of three different cinnamyl alcohols: p-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol. The formulation of lignin and the ratio of the three alcohol units change across the different types of cells and plants. Lignin compounds can be classified into two broad groups, namely, ether (C–O) linked lignins and carbon-carbon (C–C) bonded lignins, which contain β-O-4, 5–5, β-5, 4-O-5, β-1, dibenzodioxocin, and β–β linkages, of which the β-O-4 linkage is predominant. Ether and C–C linkage patterns vary with different types of substitutions among plant species [1,2].

During the past three decades, several studies have been carried out on lignin degradation using mechanical, thermal, extractive, chemical, enzymatic, and microbial approaches [1,2]. In our previous study [3], results from density functional theoretical (DFT) calculations on lignin model compounds not only were successful in predicting the bond dissociation energies (BDE) and reactivity trend of experimentally observed product selectivities, but also emphasized the roles of electron delocalization and methoxy-group effects on radical cation formation.

Here, we calculate the BDEs of the ether and C–C bond linkages in 65 unique lignin model compounds with different substituents on the arene rings and aliphatic carbons connecting the rings for which no measured bond strengths exist. DFT calculations were carried out using the M06-2X hybrid exchange-correlation functional and the 6-311++G(d,p) basis set in the gas phase. We have chosen 33 ether-linked lignin model compounds from β-O-4, α-O-4, and 4-O-5 subtypes (Fig. 1). To study representative C–C bonds (Fig. 2), we have selected 32 model compounds containing the subtypes of β-1, α-1, β-5, and 5-5 C–C linkages.

Figure 3 illustrates the ranking of the calculated BDE ranges at 298 K (corrected for zero-point energies) averaged over the respective subgroup for lignin model compounds with ether and C–C linkages considered in this work. To avoid confusion, Cα-Cβ double-bond linkages were grouped separately from the single Cα-Cβ bond linkage of β-1 model compounds. Overall, except for the 4-O-5 type, the β-O-4 and α-O-4 are found to be weaker than the C–C linkages. Ether bonds in 4-O-5 type linkages are the strongest and also exhibit the shortest bond lengths, as expected. In the Cα-Cβ linkage category, β-1 type bonds are
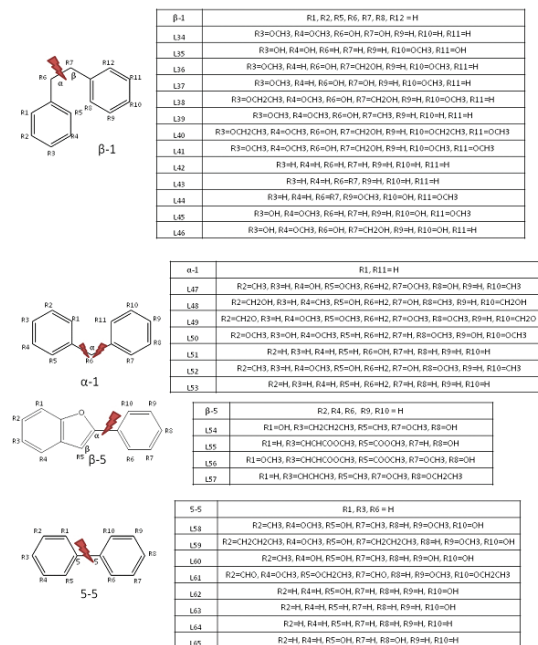
Fig. 2. Lignin model compounds (β-1, β-1, β-5, and 5-5) with carbon-carbon linkages.

the weakest, followed by the β-O-4 linkage types. It is quite likely that the fragmentation in these β-O-4 linkage types is facilitated by the weaker ether linkages instead of the Cα-Cβ cleavage. α-1 types show higher C–C BDEs compared to the β-O-4 and β-1 linkage types. Our calculations reveal that it will be difficult to cleave the β-5 and 5-5 C–C bonds as they exhibit BDEs of 110 kcal/mol or more. Therefore, fragmentation of these linkages probably requires catalysts or ring deformation reactions.

Within the α-O-4 linkages category, methoxy group substitutions at the di-ortho position in the arene ring adjacent to the ether bond give rise to the lowest BDEs, single methoxy substitutions at ortho positions result in higher BDEs, and even higher BDEs are seen in α-O-4 linked compounds when there are no substitutions in the ortho position. The weakening of ether linkages can be attributed to steric hindrance induced by the ortho-methoxy groups. In the C–C linkage category, the lowest BDEs are seen in the β-1 Cα-Cβ bond types. An interesting observation is that in β-1 model compounds, substitution of a hydroxymethyl group (CH$_2$OH) at the Cβ position results in higher BDEs, whereas substitution at the Cα position with a hydroxyl group or substitution at the Cβ position with a methyl, hydrocarbon, or hydroxyl group lowers the Cα-Cβ bond strength. Unsurprisingly, β-1 type linkages with Cα-Cβ double bonds exhibit the highest BDEs among all model compounds containing C–C bond linkages.

In summary, we have calculated BDEs of ether and C–C linkages in 65 lignin model compounds. The model compounds were selected with different linkages containing key substitutions on the arene rings and on the linkage aliphatic carbons that might significantly change the ether and C–C linkage BDEs and, hence, the homolysis selectivity. In general, it is easier to fragment the ether bond than the C–C bond. The ether bond in the α-O-4 linkage type is the weakest linkage among

the different ether lignin linkages studied in this investigation. The C–C bond of β-1 linkages has the lowest bonding strength among the lignin model compounds with C–C bond linkages. The ether linkage is substantially weakened when substituted by electron-donating methoxy groups on the adjacent arene ring. The strength of the C–C linkage is influenced by hydrocarbon, methyl, and hydroxyl group substitutions at the aliphatic carbon positions in the linkage. The reported results provide strategic avenues for delignification experimental protocols. They also offer details that could be useful in the selection and genetic design of plants having types of lignin structures that lead to optimal efficiencies for feedstock and biofuel production. Results were recently published [4].

Fig. 3. Comparison of averaged BDEs for the lignin model compounds with ether and C–C linkage groups. Standard deviations are also shown.
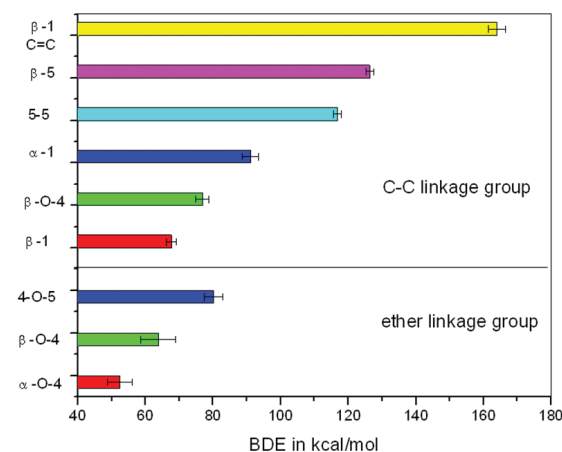
[1] Zakzeski, J. et al., *Chem Rev* **110,** 3552 (2010).
[2] Dorrestijn, E. et al., *J Anal Appl Pyrol* **54,** 153 (2000).
[3] Cho, D.W. et al., *J Org Chem* **75,** 6549 (2010).
[4] Parthasarathi, R. et al., *J Phys Chem Lett* **2,** 2660 (2011).

Exploring a New Class of Molecules Related to Cancer, Stem Cells, and Epigenetics: Long Noncoding RNAs

Irina V. Novikova, Scott P. Hennelly,
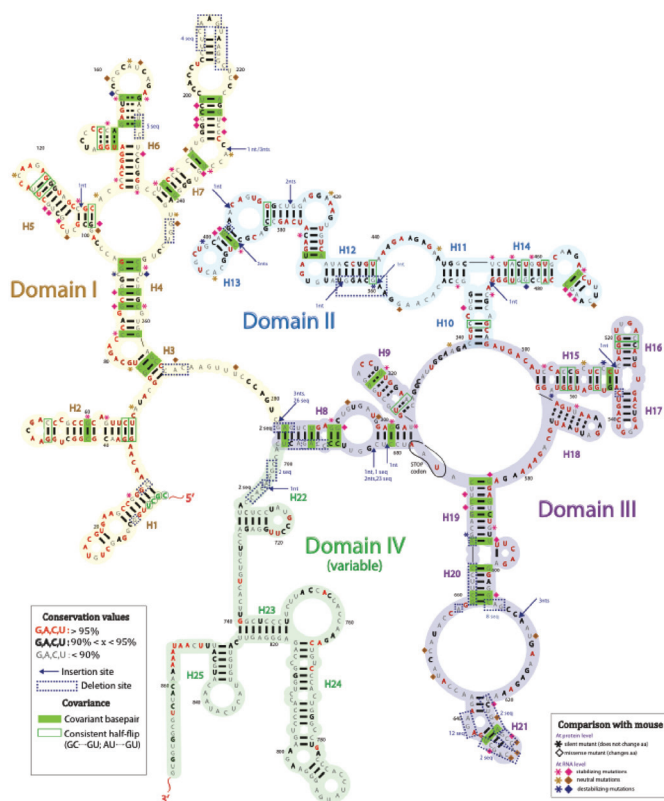Karissa Y. Sanbonmatsu, T-6

Human long noncoding RNAs (lncRNA) are a newly discovered class of biological molecules shown to play important roles in a wide range of biological functions, including hormone signaling, embryonic stem cell differentiation, development, brain function, and cancer [1–5]. In recent years it has been shown that as much as 98% of the human genome does not code for protein. In the past three years alone, approximately 50,000 lncRNAs have been discovered. As each lncRNA is typically 1,000–10,000 residues in length (1–10 kB), lncRNAs stand to occupy a substantial fraction of the human genome. lncRNAs are often associated with chromatin remodeling and other epigenetic processes. In some cases, lncRNAs recruit the polycomb repressive complex to chromatin, facilitating modification of histone proteins. Many lincRNAs are involved in stem-cell reprogramming, either regulating or being regulated by three key stem cell proteins: Nanog, Oct4, and Sox2. One of the most famous lncRNA systems is the X chromosome. In females an entire X chromosome is almost completely shut down by a suite of gigantic lncRNAs (e.g., Xist ~17 kB, Tsix ~40 kB). Here, the Xist lncRNA is overexpressed in massive quantities and covers the entire chromosome in a "chromatin coating" process.

Fig. 1. Secondary structure of the steroid receptor RNA activator lncRNA determined from chemical probing and covariance analysis. Dashed purple boxes, deletions that occurred in at least 1 of 36 vertebrate sequences. If deletion occurs in more than one species, the number of species undergoing deletion at this position is specified. Purple arrows: insertion positions. Number of nucleotides, x, incorporated at insertion site is indicated by "x nts." If insertion occurs in more than one species, the number of species undergoing insertion at this position is specified. Green-filled boxes, covariant base pairs. Green-outlined boxes, base pairs undergoing a change from a Watson-Crick base pair to a GU or UG, which do not have any instances of mismatches across all the organisms. Mutations from mouse to human tend to stabilize the RNA structure of the lncRNA.

We have produced the first secondary structure of a human long noncoding RNA ("link RNA," lncRNA) using a combination of extensive chemical probing experiments and computational sequence comparison analyses; however, the structures of these molecules remain a mystery [1-5]. As no structural studies have been performed, the following questions remain unanswered: 1) Are lncRNAs highly structured or disordered? 2) Do they contain globular sub-domains or are they organized linearly in chains of stem-loops? 3) Do lncRNAs exist in ribonucleoprotein complexes or as isolated RNAs that transiently interact with proteins? and 4) Do these molecules contain a compact core or are they more extended? We have answered the first two questions for a particular lncRNA using a combination of experimental and computational analyses [6]. In addition, very little is known about the evolution of lncRNAs. Using comparative structure analysis, we study the evolution of a lncRNA.

The subject of our study is the steroid receptor RNA activator (SRA), a lncRNA strongly associated with breast cancer [7-10]. While this molecule acts mainly as a noncoding RNA, alternative splicings of the molecule code for a protein called SRA protein, or SRAP. The steroid receptor RNA activator (874 residues) is a key player in hormone receptor regulation. This lncRNA is a nuclear co-activator and upregulates a wide variety of nuclear receptors, including the estrogen, androgen, progesterone, and thyroid hormone receptors, as well as the dosage-sensitive sex reversal (DAX-1) and steroidogenic factor (SF-1); these are key players in sex determination during development. The expression of the coding and noncoding forms of SRA occurs in different ratios for different tumor types. For this reason, many are interested in using SRA as an early onset marker for breast cancer. Understanding the molecular mechanism of SRA will help guide its usage as a marker.

Secondary, or 2D, structure in RNA is very different from 2D structure in proteins. While protein 2D structure reports on the helical versus sheet-like nature of the conformation, RNA 2D structure reports simply on Watson-Crick base pairing. However, the aggregate map of base pairing for the entire molecule yields an enormous amount of information on the architecture of the RNA, including the existence of sub-domains,
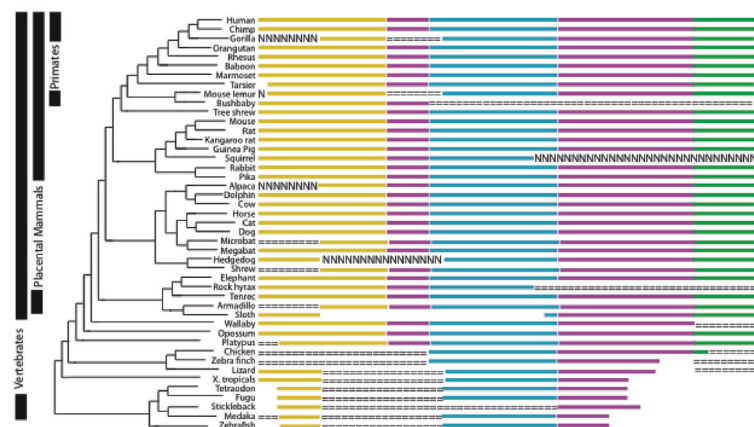
Fig. 2. Phylogenetic tree displays evolution of RNA structural domains across 45 vertebrates possessing the SRA gene. Yellow bars, domain I; cyan bars, domain II; purple/magenta bars, domain III; green bars, domain IV; "===" and "NNN" denote uncertainty in sequence alignment according to ENCODE database conventions.

the type of junction between helices, and particular bulged structural motifs. In some cases, the 2D structure alone reveals the functional mechanism itself (e.g., riboswitches). In these cases it is not necessary to solve the 3D X-ray structure to determine the mechanism of function.

The ribosome is one of the only large RNAs studied in mechanistic detail to date, so we followed in the footsteps of ribosome researchers to determine the 2D structure of SRA. We performed dimethyl sulfate chemical probing and RNase V1 digestion to locate base-paired and non-base-paired regions of the RNA. In addition, we employed newer probing techniques developed to study riboswitches. In-line probing and selective 2' hydroxyl acylation by primer extension both produce single-nucleotide resolution; they report on the RNA backbone mobility, enabling us to identify base-paired residues. We used multiple sequence alignment across 45 species to verify our helices through covariance. So, if a GC base pair in humans changes to another Watson-Crick base pair (AU, UA, or CG) in another species, this is evidence that supports our structure.

Our results show the lncRNA to be highly structured (Fig. 1) and organized into four sub-domains. In all, we identified 25 RNA helices, 16 terminal loops, 15 internal loops, and 5 junction regions. Our structure is consistent with previous in vivo site-directed mutagenesis and deletion studies. Our structure has several functional implications. It is currently not known whether SRA co-activates hormone receptors by escorting them through the nuclear membrane or by acting as a structural scaffold for chromatin in the transcription complex. We have shown that SRA is highly structured, suggesting it is possible that SRA provides structural scaffolding for the transcription complex. In addition, it was previously suggested that H13 (or Str 7) is essential

for hormone-receptor binding. However, a variety of other proteins interact with SRA, including DAX-1, SF-1, deadbox proteins P68/P72, Myo-D, SLIRP, SHARP, Pus1, and Pus3. We find the three-way junction branching helices H15, H16, and H17 to be much more highly conserved than H13. Thus, this junction region may be a second protein-binding site.

Another interesting aspect of the SRA gene is that it functions as an RNA in some cases but as a protein in others. It has been proposed that at some point in evolution, the noncoding Xist lncRNA (responsible for X-chromosome inactivation) might have originated gradually, allowing for a period of time where noncoding and coding isoforms of the gene coexisted. The coding and noncoding isoforms of SRA also originated from the same gene, suggesting that this lncRNA might be a rare and unique capture of this stage of evolution. To examine this more carefully, we studied the evolution of SRA (Fig. 2). Since the 3D structure of the mouse SRAP has been solved by nuclear magnetic resonance, we had the unique opportunity to study the co-evolution of RNA structure and protein structure. We found rapid evolutionary stabilization of the RNA structure. The vast majority of mutations from mouse to human act to stabilize RNA helices. In contrast, most mutations of protein sequence occur in linking regions between helices and do not appear to be obviously stabilizing. In addition, several frame-disrupting mutations occur from mouse to human in conserved regions, suggesting that evolutionary pressure preserves the RNA structural core rather than its translational product.

[1] Ulitsky, I. et al., *Cell* **147,** 1537 (2011).
[2] Guttman, M. et al., *Nature* **477,** 295 (2011).
[3] Huarte, M. et al., *Cell* 1**42,** 409 (2010).
[4] Gupta, R.A. et al., *Nature* **464,** 1071 (2010).
[5] Ponting, C.P. et al., *Cell* **136,** 629 (2009).
[6] Novikova, I.V. et al., *Nucleic Acids Res,* in press (2012).
[7] Lanz, R.B. et al., *Cell* **97,** 17 (1999).
[8] Chooniedass-Kothari, S. et al., *FEBS Lett* **566,** 43 (2004).
[9]Hube, F. et al., *DNA Cell Biol* **25,** 418 (2006).
[10] Yan, Y. et al., *Breast Cancer Res* **11,** R67 (2009).
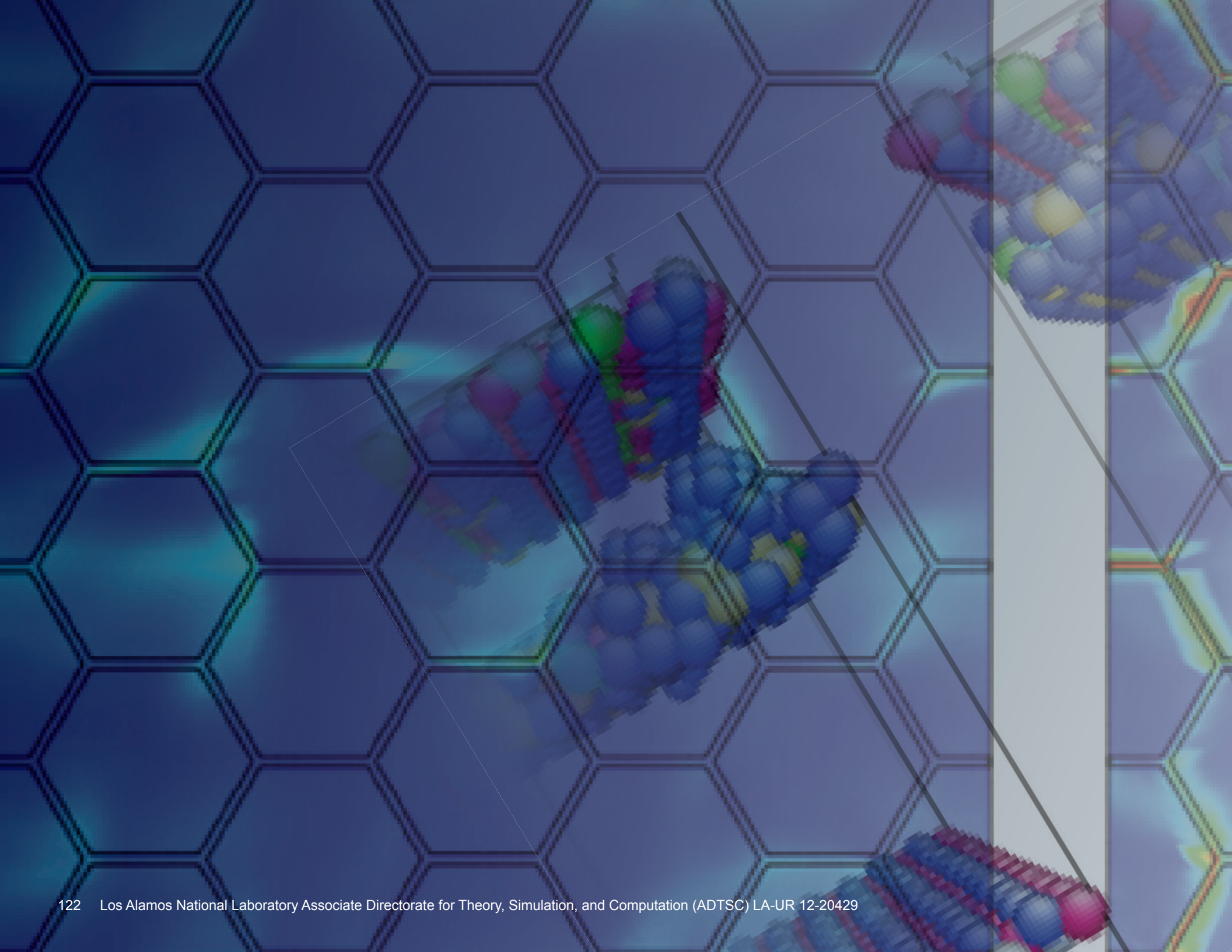
Materials Science

As a result of our national security mission and over 60 years of scientific research, LANL is a world-class material science research and development institution. LANL scientists conduct fundamental research into the state and nature of materials, provide new theories for materials behavior, and create and provide new constitutive properties models and parameters for a wide variety of simulation tools that are in use or under development around the world. Not surprisingly, much of the work at LANL deals with materials in extreme conditions–highly stressed, shocked, and in high-radiation field environments. This work in particular supports the grand vision of the Laboratory to design and construct a world-class facility called MaRIE (Matter-Radiation Interactions in Extremes) for the prediction and control of material performance under dynamic and harsh conditions. The articles included in this section describe work at multiple scales–from atomistic to meso–involving the use of quantum mechanical and molecular dynamics methods, as well as the study of specific material properties and the derivation of new and more accurate material models under various conditions. This broad range of physical modeling approaches, mathematical and numerical algorithms, and multiple length and time scales is illustrative of both the breadth and depth of LANL's capabilities in this arena.

A Single-Crystal Model for High-strain-rate Applications

Francis L. Addessio, Curt A. Bronkhorst, T-3;
Donald W. Brown, Ellen K. Cerreta, MST-8;
Xiangdong Ding, T-4; Todd L. Graves, CCS-6;
Turab Lookman, T-4; Paulo A. Rigg, WX-9;
Michael E. Wall, CNLS

A thermodynamic framework is being used to model single-crystals for the extreme conditions of high-strain rate and high pressure. The effects of nonlinear elasticity, solid-solid phase transformations, plastic slip and twinning, and damage are included in the approach. The single-crystal model will be implemented into a structural analysis and used to pursue a polycrystal response. Coarse-graining or homogenization techniques will be considered in an effort to generate macro-mechanical models. The effort is closely coupled with experimental, microscopy, and diffraction studies.
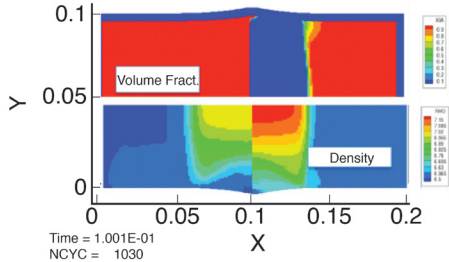
redictive material models are necessary for the design and life extension of structures as well as the interpretation of subscale experimental investigations. Accurate models are required to provide deformation characteristics for a diverse range of loading scenarios. Predictive models are especially important for extreme conditions where experimental investigations cannot be conducted. To simulate component performance, increased burden is placed on modeling the effects of manufacturing processes on materials and on predicting their design behavior. In the past, uncertainty was mitigated with conservative designs. The macro-mechanical response of materials has its basis in the microstructure of the material. That is, deformation processes are dictated by mechanisms at the subgrain scale. The single-crystal or meso-mechanical length scale (the order of 50 μm) provides an intermediate length and time scale, which bridges the scales that are modeled using atomistic and macro-mechanical approaches.

The physics of solid-solid phase transformations has been addressed at many length scales. Molecular dynamics (MD) simulations have examined the effects of shear, hysteresis, and high-strain rates on the transformation process. Micro-mechanical models, which incorporate the details of nucleation, interface motion, and twin growth, have demonstrated the ability to describe the fundamental phenomena inherent to the transformation processes [1,2]. Meso-mechanical models have explored the details of phase transformations at the single-crystal scale [3-5]. However, few single-crystal approaches address the conditions, which are inherent to high-pressure and strain-rate phenomena [6,7]. Macro-mechanical approaches are necessary for addressing the deformation process of engineering structures.
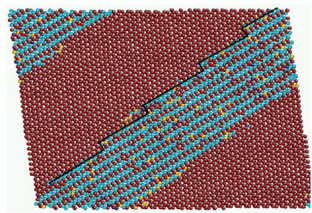
Macro-mechanical models do not, in general, address issues related to meta-stability, hysteresis, retained high-pressure phases, and the effects of shear on the transformation processes.

A model is being developed for the large deformation of single crystals under the conditions of high pressure and high-strain rate. The effects of deformational and transformational twinning, plastic slip, phase transformations, and damage are being considered. The model is not intended to resolve individual transformation interfaces or domains. Instead, volume fractions of the constituents within a representative volume of the material are modeled. Each constituent that is represented by a volume fraction is assumed to have the same material properties, such as lattice orientations, strain, and temperature. The constituents are also assumed to rotate together. A number of details are omitted in the model, including the nucleation of dislocations and damage. A thermodynamic framework, which is developed in terms of free energies, is used. In the model, the total deformation gradient (F) is divided into contributions due to elasticity (F^e), plasticity and damage (F^p), and the transformation process (F^ϕ):

$$F = F^e \, F^p \, F^\phi \qquad (1)$$

The velocity gradient (L^p) of the plastic component is obtained using sums over slip (s) and deformation twin (t) systems:

$$L^p = \dot{F}^p F^{p-1} = \sum_s \dot{\gamma}^s \, (\vec{s}^{\,s} \otimes \vec{n}^{\,s}) + \sum_t \gamma^t \, (\vec{s}^{\,t} \otimes \vec{n}^{\,t}) \qquad (2)$$

In Eq. (2), $\gamma, \vec{n}, and \ \vec{s}$ are the deformation resistance, slip or twin plane normal, and shear directions. The deformation gradient for the transformation process is obtained using sums over the transformation systems (ϕ):

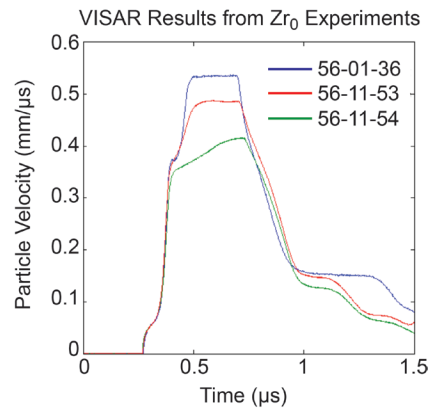$$\dot{F}^\phi = \sum_\phi \dot{\xi}^\phi \gamma^\phi (b^\phi \otimes m^\phi) \qquad (3)$$

Fig. 1. Simulation of a plate impact experiment. Upper contour of volume fraction and lower contour of density (g/cm**3).

Time = 1.001E-01
NCYC = 1030

Fig. 2. Molecular dynamics simulation of shear on a Zr sample.

Fig. 3. Particle velocity measurements for recent plate impact experiments.
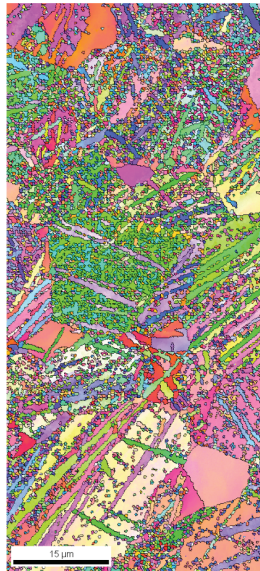
In Eq. (3) ξ^ϕ, γ^ϕ, $and\ \bar{m}^\phi$ are the mass fractions, shape strain, and normal to the transformation plane. Computational solutions of the resulting system of equations are obtained using an extension to existing approaches for single-crystal plasticity [8].

The driving force for phase transformations includes terms due to mechanical work, thermal energy, surface energy, and defect energy of the constituents. These forces will be extended to address the high strain-rate applications that are of interest. Although a general model formulation is being pursued, the α (hexagonal close packed) to ω (hexagonal) phase transformation in titanium (Ti) or zirconium (Zr) is being considered because of the plethora of experimental and atomistic information that is available for these elements. The model will allow the ability to include the effects of hysteresis, retained high-pressure phase, and kinetics of the transformation process.

A plate impact simulation of Zr flyer (0.0 < x < 0.1 cm) and target (0.1 < x < 0.2 cm) plates is provided in Fig. 1, using a preliminary version of the proposed model. Each computational cell within the target plate of the simulation is assumed to be a single crystal. MD simulations (Fig. 2) have been used to explore the effects of shear and thermal cycling on the transformation process of Zr. Furthermore, MD simulations and experiments have been instrumental in defining material parameters for the single-crystal model. Polycrystal simulations will be used to pursue averaged quantities, which are necessary to explore improved macro-mechanical models for transformation processes. Coarse graining techniques also will be considered as a means to generate macro-mechanical models.
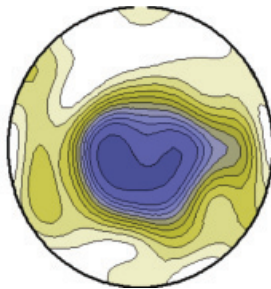
This effort is coupled to small-scale experimental investigations. Plate impact experiments have pursued the effects of peak pressure and temperature on the transformation process and meta-stability of Zr. Recently, two experiments were conducted at peak pressures of 8.0 GPa and 10.5 GPa (Fig. 3). The particle velocity response of the 8.0 GPa experiment

Fig. 4. Microscopy and diffraction results for a plate impact sample of Zr.

indicated a sluggish transformation process relative to higher peak pressure experiments. Furthermore, whereas ~ 35% retained high-pressure (ω) phase was observed in past experiments, ~63% and ~82% retained high-pressure phase were observed in the 8.0 GPa and 10.5 GPa experiments, respectively. Diffraction investigations also suggest the presence of another crystalline structure. This result could provide insight into the potential pathway for the high-pressure transformation. Whether the transformation process is completed during the initial compression of the sample still is unknown. These results along with past torsion experiments, which resulted in ~95% retained high-pressure (ω) phase, suggest the importance of shear on the transformation process. Currently, the effects of shear and rate are not included in the existing phase transformation models that are used in engineering analyses. Future efforts will focus on microscopy (Fig. 4) and diffraction investigations of the plate impact samples. Information regarding the texture, transformation systems, and mechanical properties of the high-pressure phase will be considered.

[1] Abeyaratne, R. and J.K. Knowles, *Arch Rational Mech Anal* **114,** 119 (1991).
[2] Ball, J.M. and R.D. James, *Arch Rational Mech Anal* **100,** 13 (1987).
[3] Turteltaub, S. and A.S.J. Suiker, *Int J Solid Struct* **43,** 4509 (2006).
[4] Suiker, A.S.J. and S. Turteltaub, *Int J Numer Meth Eng* **63,** 1655 (2005).
[5] Tjahjanto, D.D. et al., *Model Simulat Mater Sci Eng* **14,** 617 (2006).
[6] Barton, N.R. et al., *Model Simulat Mater Sci Eng* **13,** 707 (2005).
[7] Becker, R., *Int J Plast* **20,** 1983 (2004).
[8] Kalidindi, S.R. et al., *J Mech Phys Solid* **40,** 537 (1992).

A Material Model for High-strain-rate Applications, Including Phase Transformations, Material Anisotropy, and Damage

Abigail Hunter, XCP-1; Thomas R. Canfield,
Francis L. Addessio, T-3

A thermodynamic framework has been used to develop a macro-mechanical model for high-strain-rate deformations. The model includes the effect of nonlinear elasticity (an equation of state), solid-solid phase transformations, plasticity, and damage. Phase transformations are addressed using either a free-energy or an equilibrium-phase diagram. An anisotropic inelastic potential is used to model the combined effects of plasticity and ductile damage. The model is being implemented into an engineering design computational framework and validated using small-scale experimental data.

There are numerous applications, which involve high-strain rates and high pressures. Impact, penetration, and weapons performance scenarios are examples of representative applications. The physical processes, which are encountered during high-rate deformations, may include nonlinear elasticity, plasticity, phase transformations, and damage. Nonlinear elasticity or equations of state are necessary to accurately model the material response to shock loading conditions. Coupled effects, including the ability of plasticity to inhibit a material from transforming back to a parent phase, for example, also represent important effects, which must be modeled. As larger stress states are encountered, damage nucleation, growth, and coalescence must be considered. These phenomena have their basis in the evolution of the material microstructure. However, macro-mechanical models are still necessary to address engineering applications.
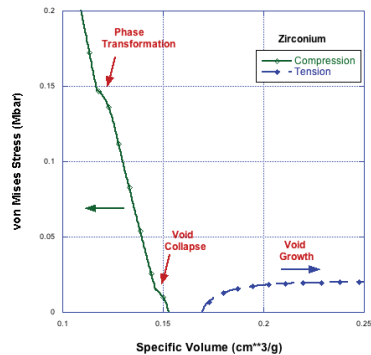
A macro-mechanical model has been developed that is based on thermodynamic considerations. To address phase transformations, the approach may utilize free energies for each phase or an equilibrium-phase diagram, depending on the availability of data. An anisotropic yield surface for the combined phenomena of plasticity and damage is considered to address the effects of material texture, rate dependence, and ductile failure. The thermodynamic framework, in general, relies on both elastic (free energy) and inelastic potentials. The total specific Helmholtz free energy for a mixture of phases is provided as the mass fraction average of the free energies of each of the k-constituents [$\psi^k(\bar{\bar{\varepsilon}}^e, T, D, \varepsilon^p)$] and a component due to mixing (ψ^{mix}):

$$\psi(\bar{\bar{\varepsilon}}^e, T, D, \varepsilon^p, \bar{\bar{\varepsilon}}^i, m) = \sum_k m_k \psi_k(\bar{\bar{\varepsilon}}^e, T, D) + \psi^{mix}(\bar{\bar{\varepsilon}}^i, m) \quad (1)$$

In equation (1), m_k is the mass fraction of each constituent, ψ_k is the free energy of the kth constituent, $\bar{\bar{\varepsilon}}^e$ is the elastic strain, T is the temperature, D is a measure of material damage, and $\bar{\bar{\varepsilon}}^e$ is the inelastic strain. Also, ψ^{mix} is the free energy, which is a consequence of mixing of the phases. Example free energies may be found in the literature [1-3]. Substitution of the mixture free-energy [equation (1)] into the Clausius-Duhem inequality for the dissipation rate results in expressions for the generalized thermodynamics forces including the stress, entropy, plasticity (hardening), damage, transformation strain, and transformation kinetics [3].

Two inelastic potentials, related to the combined effects of plasticity and damage [$\phi^p(\bar{\bar{\Sigma}}, \bar{\varpi}^p, \bar{\varpi}^d)$] and phase transformations [$\phi^t(\Pi) = 0$], are postulated. The appropriate inelastic parameters then may be defined in terms of these potentials [3]:

$$d\bar{\bar{\varepsilon}}^p = \frac{\partial \phi^p}{\partial \bar{\bar{\Sigma}}} d\lambda^p, \qquad d\varepsilon^p = \frac{\partial \phi^p}{\partial \bar{\varpi}^p} d\lambda^p$$

$$dD = \frac{\partial \phi^p}{\partial \bar{\varpi}^d} d\lambda^p, \qquad dm = \frac{\partial \phi^t}{\partial \Pi} d\lambda^t \qquad (2)$$

In equation (2), the thermodynamics variables are $\bar{\bar{\Sigma}} \equiv \bar{\bar{\sigma}} - \bar{\bar{\mu}}^i$ and $\Pi \equiv \bar{\bar{\Sigma}} : \bar{\bar{\Lambda}} - \mu^m$. The parameters $d\lambda^p$ and $d\lambda^t$ are the Lagrange multipliers for plasticity and phase transformations, which may be obtained from the consistency conditions ($d\phi^p = 0$ and $d\phi^t = 0$). Extensions of the Gurson surface [4] have been considered for an ellipsoidal void embedded within a perfectly plastic, anisotropic matrix [5]. In the current model development, it is assumed that the voids remain spheroidal. The flow surface can be written with respect to a Cartesian coordinate system [5]
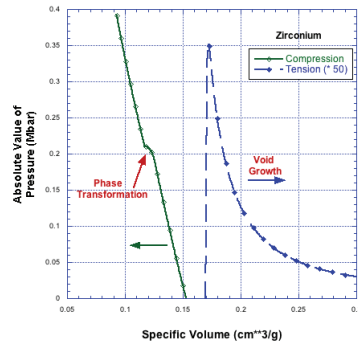

Fig. 1. Pressure versus specific volume for uniaxial compression and tension of zirconium.

Fig. 2. Von Mises stress versus specific volume for uniaxial compression and tension of zirconium.

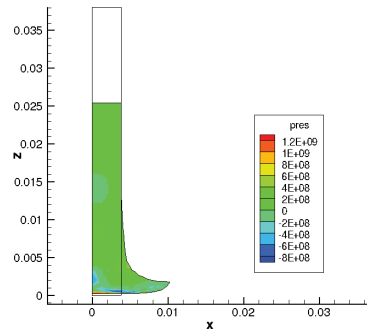*Fig. 3. A representative simulation of a Taylor impact experiment showing contours of pressure at t= 60 ms. The wire outline provides the original shape of the projectile.*

associated with the axis of orthotropy for the matrix material:

$$\phi^p(P,\overline{\overline{s}},\phi) = \frac{3}{2}\overline{\overline{\sigma}} : \overline{\overline{\overline{M}}} : \overline{\overline{\sigma}}$$

$$- Y_s^2\left[1 + (q_1\phi)^2 - 2q_1\phi\cosh\left(\frac{q_2\kappa P}{Y_0}\right)\right] = 0 \qquad (3)$$

In equation (3), $\overline{\overline{\overline{M}}}$ is the anisotropy tensor for the matrix material, ϕ is the porosity, $\overline{\overline{\sigma}}$ is the Cauchy stress, P is the pressure, and κ is a material constant that is dependent on the material anisotropy. Strain rate dependent models may be used for the yield function (Y_s) of the solid or matrix material. The effective plastic strain of the matrix (solid) constituent (ε_s^p) is used as the hardening parameter in the model for the yield function. The matrix plastic strain is obtained from the equality between the plastic work of the composite and matrix materials [6]. The inelastic potential for the phase transformations is written in terms of the differences [2] of the free energies ($\dot{m}_k = f_k(\Delta g_{sk})$, where $g_{sk} = \psi_{sk} - v_{sk}\overline{\overline{\sigma}}_{sk} : \overline{\overline{\varepsilon}}_{sk}^e$ is the Gibb's free energy of each phase). If free energies are not available, then the kinetics may be obtained from an equilibrium phase diagram [7].

Finally, the total strain rate is decomposed into the elastic ($\dot{e}^e$), plastic ($\dot{e}^p$), and transformation ($\dot{e}^t$) rates. The plastic strain rate is obtained directly from the plastic potential [equation (2)]. Functional forms for the transformation strain rate may be found in the literature [3]. The evolution of porosity, in general, is composed of contributions due to void nucleation, growth, and coalescence. Classical approaches model the porosity growth [4,6] component as directly related to the change in the volumetric plastic strain [$d\varepsilon^p = tr(d\overline{\overline{\varepsilon}}^p)$] of the composite material $d\varphi = (1-\phi)d\varepsilon^p$.

A novel numerical technique [8] has been used to solve the resulting equations for the shear stress ($\tau = 3\overline{\overline{s}} : \overline{\overline{s}} / 2$), pressure ($P$), porosity ($\phi$), and plastic strain, once the kinetics of the transformation process have been addressed. An overstress approach is being considered to address the issues related to localization and the ensuing ill-possedness problems.

Single-cell simulations (Figs. 1 and 2) using the model demonstrate the ability of the model to capture porosity crush-up and phase transformations for a uniaxial compressive condition as well as the growth of damage for a uniaxial tensile condition. The model is being implemented into a finite-difference computational analysis. Experimental data will be used to validate the model. For example, plate impact and Taylor impact (Fig. 3) simulations will be used in comparisons with experimental data to explore the ability of the model to accurately capture material anisotropy, rate dependence, phase transformations, and damage.

[1] Lemaitre, J. and J.-L. Chabouche, *Mechanics of Solid Materials,* Cambridge University Press, New York (1990).
[2] Greeff, C.W., *Model Simulat Mater Sci Eng* **13,** 1015 (2005).
[3] Boyd, J.G. and D.C. Lagoudas, *Int J Plast* **12,** 805 (1996).
[4] Gurson, A.L., *J Eng Mater Tech* **99,** 2 (1977).
[5] Monchiet, V. et al., *Int J Plast* **24,** 1158 (2008).
[6] Tvergaard, V. and A. Needleman, *J Mech Phys Solid* **34,** 213 (1986).
[7] Addessio, F.L. et al., *J Appl Phys* **93,** 0021 (2003).
[8] Maudlin, P.J. et al., "Material Anisotropy," unpublished report (2008).

# Possible Self-healing in Tungsten under Fusion Reactor Conditions

Valery Borovikov, T-1; Xian-Zhu Tang, T-5;
Danny Perez, T-1; Xian-Ming Bai, INL;
Blas P. Uberuaga, MST-8; Arthur F. Voter, T-1

Radiation damage by fast neutrons due to fission and fusion can significantly degrade material properties. In a fusion reactor, long-lasting radiation-induced defects such as vacancies, vacancy clusters, and voids introduce additional nuclear safety complication in terms of trap sites for excessive tritium retention. This is especially true for the plasma-facing components of a fusion reactor like ITER, where extreme thermal stress is also present. Here we use molecular dynamics simulations to elucidate a self-healing mechanism in which the large thermal stress can facilitate the recombination of the neutron-collision-cascade-induced vacancies and interstitials through coupled grain boundary (GB) motion in body-centered-cubic (bcc) tungsten under fusion reactor conditions. Specifically, our simulations reveal that for a number of tungsten GBs, absorbing the fast-moving interstitials can help activate coupled GB motion at reduced mechanical stress; the migrating GB then sweeps up the less-mobile vacancies, facilitating vacancy-interstitial recombination inside the GB.
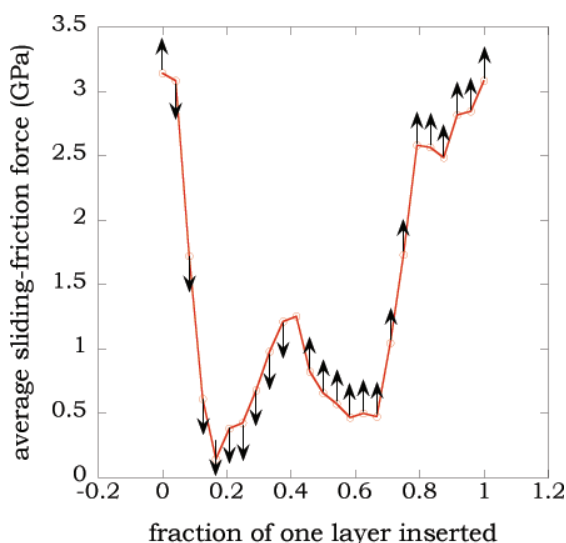
A critical problem in developing fusion as a future energy source is designing materials that will tolerate the harsh radiation environment in a tokamak reactor. The challenges for tungsten (W), or any other plasma-facing component (PFC) in a fusion reactor, are compounded by the unusual combination of extensive radiation damage and strong mechanical stress. The neutrons produced in a fusion reactor can damage any structural material. The primary damage takes the form of interstitial atoms and vacancies. These defects and their aggregates not only affect the mechanical, thermal, and electrical properties [1,2] of the PFC, but also introduce a number of additional nuclear safety concerns [3-5]. Healing of the material occurs when the vacancies and interstitials recombine and annihilate. In metals, however, the interstitials are highly mobile. They quickly diffuse to the nearest grain boundaries (GB) or surface, leaving the vacancies behind in the bulk of the material. Structural materials must be designed such that they tolerate the harsh radiation environment in a fusion reactor.

We have been studying the influence of radiation-induced damage on the GB sliding process in body-centered-cubic (bcc) W, employing atomistic computer simulations. For a number of GBs, we found the surprising result that introducing interstitials or vacancies into the GB can reduce the average sliding-friction force under shear by more than an order of magnitude (see Fig. 1). Moreover, because these GBs typically shear in a well-known coupled way (linked sliding and normal migration) [6], we propose the following new self-healing mechanism for W under irradiation conditions. A collision cascade produces vacancies and highly mobile interstitials; the diffusing interstitials find, and are trapped at, a nearby GB. The interstitial-loaded GB is now so easy to shear that internal stresses in the crystal may set it in motion, sweeping up the vacancies close to the cascade center. Thus, under the right conditions, the very cascade that introduces damage into the material can initiate a mechanism that promotes healing of that damage. This mechanism is crucially dependent on the presence of a large mechanical stress, and the extreme stress conditions in a fusion reactor multiple gigapascal (GPa) appear to meet this requirement, at least during the plasma instabilities [7]. In our simulations we have also observed that the direction of normal GB motion switches to the opposite direction for many cases when the defects are introduced into the GB (e.g., see Fig. 1).

The self-healing phenomena under fusion reactor conditions can be illustrated by molecular dynamics simulations of a nanometer-scale W system with a $\Sigma5(013)[100]$ symmetric-tilt grain boundary. In the beginning of the simulation we introduce the interstitials into the GB and the vacancies into the upper grain (see Fig. 2). The sample is maintained at T = 1000 K, which is the nominal operating temperature

Fig. 1. Average sliding-friction force, under an applied shear strain, as a function of number of interstitials introduced into $\Sigma5(013)[100]$ grain boundary. The direction of normal grain boundary motion is indicated by the arrows. The points without arrows correspond to pure sliding.
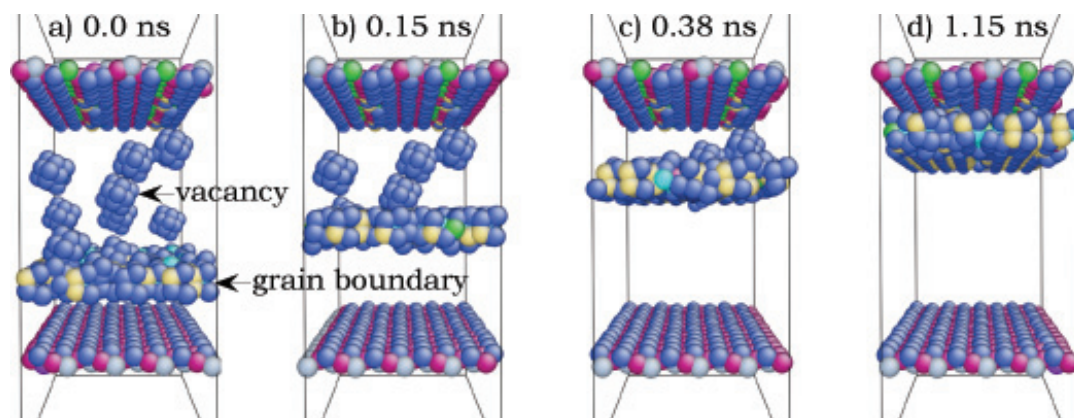
Fig. 2. Demonstration of the self-healing mechanism. A threshold stress for activation of the grain boundary motion is decreased with introduction of the interstitials into the grain boundary. The grain boundary migrates under an applied stress. It sweeps vacancies on its way up, leaving behind clean material without the defects. The top and bottom layers correspond to the free surfaces. The atoms are colored according to their coordination number, and bulk atoms are not shown.

accessible mechanical stress. This sweeping mechanism illustrates the complexities associated with the atomic-scale interactions of point defects with GBs that may be crucial for understanding radiation damage evolution. It may contribute to the design of new structural materials for service in extreme environments, such as fusion reactors.

of the PFC in a fusion reactor, and a shear stress of $\sim$2 Gpa is applied in the horizontal direction. Under these conditions, the $\Sigma 5(013)[100]$ GB performs a coupled–normal migration plus parallel sliding–motion. The upward normal migration allows the GB to sweep up the bulk vacancies, which are largely immobile on the GB migration time scale. The healing of the radiation damage is evident as the sweeping GB leaves behind clean bulk W.

Experimentally, under irradiation, GBs have been observed to move in polycrystalline materials, as in electron-irradiated steel [8], and both void denuded zones [9] and segregant concentration profiles [10] have been observed to be asymmetrically distributed around GBs, consistent with GB motion. The observed reversal of the direction of GB motion as grains are traversed [8] might be reinterpreted in light of our results as a sign of a changing defect content.

Because GBs tend to act as efficient sinks for interstitials that are produced in collision cascades, and because interstitial loading enhances GB mobility as well as sometimes altering the coupling properties, we conclude that under the extreme stress conditions in a fusion reactor GB sweeping may contribute to radiation damage healing. Moreover, this mechanism may produce an effect even in systems where interstitials do not affect GB mobility, if the GB motion is coupled under

[1] De La Rubia, T.D. et al., *Nature* **406,** 871 (2000).
[2] Bacon, D.J. and Yu.N. Osetsky, *J Miner Met Mater Soc* **59,** 40 (2007).
[3] Johnson, D.F. and E.A. Carter, *J Mater Res* **25,** 315 (2010).
[4] Xu, Q. et al., *J Nucl Mater* **367,** 806 (2007).
[5] Brooks, J.N. et al., *Nucl Fusion* **49,** 035007 (2009).
[6] Cahn, J.W. et al., *Acta Mater* **54,** 4953 (2006).
[7] Pestchanyi, S. and J. Linke, *Fusion Eng Des* **82,** 1657 (2007).
[8] Singh, B.N. and T. Leffers, *Scripta Metall Mater* **8,** 549 (1974).
[9] Shaikh, M.A., *J Nucl Mater* **187,** 303 (1992).
[10] Sakaguchi, N. et al., *J Mater Sci* **40,** 889 (2005).

Coupling High-Energy Ignition with Mechanical Behavior

Bradford E. Clements, T-1; Bryan Henson,
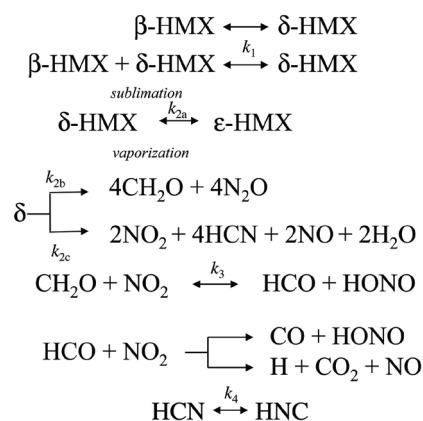Laura B. Smilowitz, C-PCS

To successfully model important National Nuclear Security Administration (NNSA) issues such as high explosives (HE) handling and safety, performance and surety, and survivability to a hostile threat, it is necessary to couple the dynamic mechanical behavior of an explosive with its thermal ignition behavior. We have put forth a promising analysis to accomplish this by using T-Division-developed explosive mechanical models with a thermal ignition model developed in the LANL C-PCS group.

The transition of an unreacted energetic material to a full-up detonated state can be classified into three general categories: 1) shock-to-detonation (SDT), 2) deflagration-to-detonation (DDT), and 3) anomalous-to-detonation (XDT) transition. SDT is the only well-understood process out of the three. SDT is prompt detonation, with typical build-up to detonation times being on the order of a few microseconds. For these short times, the underlying mechanism required to cause a heterogeneous explosive to reach a temperature high enough for a prompt detonation (typically in the range of one to two thousand Kelvin) is energy localization in the form of so-called hotspots. Because the physics of hotspots has been intensely studied, the underlying mechanisms at work in SDT are also believed to be understood. The role of an energetic material's mechanical behavior is believed to be of secondary importance in SDT, although it is known that the amount of damage, contained inclusions, binder concentration, etc., all quantitatively effect SDT to some degree. DDT and XDT are much less well understood, which is unfortunate because these two categories encompass very important NNSA issues: high explosive handling and safety, performance and surety, and survivability to a hostile threat. In DDT and XDT, mechanical properties have a dominating role and will influence the outcome of an initiated chemical reaction—that is, will the reaction extinguish, build up to detonation, or something else in between? Moreover the time for thermal runaway may be very long, say tens of microseconds up to hours. Clearly, to model DDT and XDT there is a need to couple a thermal mechanical model with a thermal ignition model, and that is the focus of the work described here.

To understand the relevant physics that must be captured by a mechanical constitutive theory, imagine a low velocity impact of a high explosive (HE) by a thrown metal fragment. By low velocity, it is meant that the impact is sub-SDT. The HEs of interest in this work are solid heterogeneous explosives, and for LANL the two explosives PBX 9501 and PBX 9502 are preeminent. These are both plastic-bonded explosives, having a high concentration of explosive grains embedded in a polymer-binding matrix. As the incoming fragment collides and penetrates the HE, considerable heating will take place from the mechanical work being done on the explosive at the expense of the fragment's kinetic energy. Adiabatic heating will be one source of temperature rise. This arises from the volumetric compression of the HE ahead of the moving fragment. The compressed HE will heat according to its equation of state (EOS). As the velocity of impact is reduced, inelastic work done on the HE grows in importance. Inelastic heat sources include plastic heating at the process zone of a growing crack in the HE, frictional heating of growing cracks in the presence of a hydrostatic pressure, plastic flow occurring in the explosive grains and the binder, and viscoelastic work done by the polymeric binder. The accumulation of all these heat sources will result in a temperature rise in the impacted explosive. It is this temperature that is fed into the thermal ignition model, and for that the present work invokes the Henson-Smilowitz thermal ignition model.

The Henson-Smilowitz thermal ignition model is physically based and has been parameterized to a wide set of independent explosive initiation experiments. Here we will focus on PBX 9501, which uses octagen (HMX) as its explosive component. Figure 1 shows a schematic of the important decomposition steps in HMX, beginning with the beta-delta solid-solid first-order phase transformation and concluding with a

Fig. 1. Schematic of decomposition steps of HMX in the Henson-Smilowitz thermal ignition model.

$$\beta\text{-HMX} \longleftrightarrow \delta\text{-HMX}$$
$$\beta\text{-HMX} + \delta\text{-HMX} \overset{k_1}{\longleftrightarrow} \delta\text{-HMX}$$
$$\delta\text{-HMX} \overset{k_{2a}}{\underset{sublimation}{\longrightarrow}} \epsilon\text{-HMX}$$

vaporization

$$\delta \overset{k_{2b}}{\longrightarrow} 4CH_2O + 4N_2O$$
$$\overset{k_{2c}}{\longrightarrow} 2NO_2 + 4HCN + 2NO + 2H_2O$$
$$CH_2O + NO_2 \overset{k_3}{\longleftrightarrow} HCO + HONO$$
$$HCO + NO_2 \longrightarrow CO + HONO$$
$$\longrightarrow H + CO_2 + NO$$
$$HCN \overset{k_4}{\longleftrightarrow} HNC$$
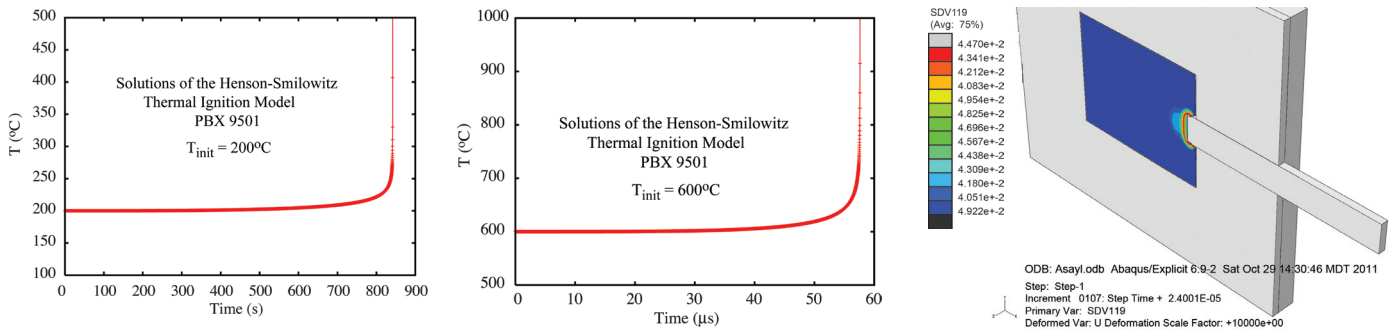
Fig. 2. Solution of the Henson-Smilowitz thermal ignition model for 200ºC (left) and 600ºC (right).
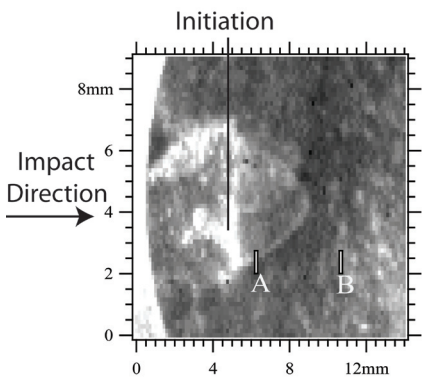

Fig. 3. Temperature profile from a simulation of the impact of PBX 9501 (left), and the corresponding experiment (right). In the left figure, red in the color scheme corresponds to about 600ºC.

gaseous mixture of decomposition molecules. The Henson-Smilowitz decomposition steps capture the chemistry of both the dark- and bright-zone burning important in HMX decomposition. Each species occurring in Fig. 1 evolves according to a differential equation (DE) for that species concentration in terms of the other decomposition products. Altogether there are six coupled DEs and two DEs describing the temperature and pressure evolution. The coefficients of the DEs are temperature-dependent Arrhenius exponentials, and this is where the temperature from the mechanics described above is first implemented. As the temperature rise from the chemical reaction begins to exceed that from the mechanical work, it is that temperature which is supplied to the Arrhenius exponentials.

Figure 2 is the solution for the Henson-Smilowitz model for two initial temperatures. It is clear that if the mechanical work can exceed 600ºC, then in a low velocity impact lasting 50 to 100 μs, run away behavior is a likely outcome. Finally, Fig. 3 shows the results of a simulation where a thermo-mechanical constitutive model, having all the properties described above, is fully coupled to the Henson-Smilowitz thermal ignition model. In the experiment and the simulation a rectangular steel beam impacts PBX 9501 at a speed of around 100 m/s. The HE simulated temperature near the impactor exceeds 600ºC and chemical reaction has begun, similar to the experiment shown in Fig. 3.

Taylor Impact Tests and Simulations of Plastic-Bonded Explosives

Bradford E. Clements, T-1; Darla G. Thompson, WX-7;
Darby J. Luscher, T-3; Racci DeLuca, WX-7

Taylor impact tests have been conducted on plastic-bonded explosives to characterize the stress state of these materials as they impact anvil surfaces at speeds between 80 and 200 m/s. Using a high-speed camera, specimen images are obtained as they undergo substantial deformation, including fragmentation. PBX 9501 and PBXN-9 are investigated. By performing finite element simulations, the Taylor test has proved to be a valuable tool for validating LANL explosives models in situations undergoing extreme damage and fragmentation.

Taylor impact tests were previously conducted on plastic-bonded explosives (PBX) to characterize the stress state of these materials as they impact smooth flat anvil surfaces at speeds on the order of 100 m/s [1]. The focus of that work was on the explosive PBXN-109. In 2003, Liu and Ellis performed Taylor tests on PBX 9501 up to speeds of 115 m/s, capturing impact images using high-speed photography [2]. The goal was to discover the threshold velocity for the initiation of the explosive. No threshold was observed. In the work presented here we extended these tests to velocities in excess of 200 m/s. We used a high-speed camera to obtain specimen images as they undergo substantial deformation, including fragmentation. PBX 9501 and PBXN-9 were investigated. While no chemical reaction was observed at these velocities, the Taylor test proved to be a valuable tool for validating LANL PBX models in situations undergoing extreme damage and fragmentation.
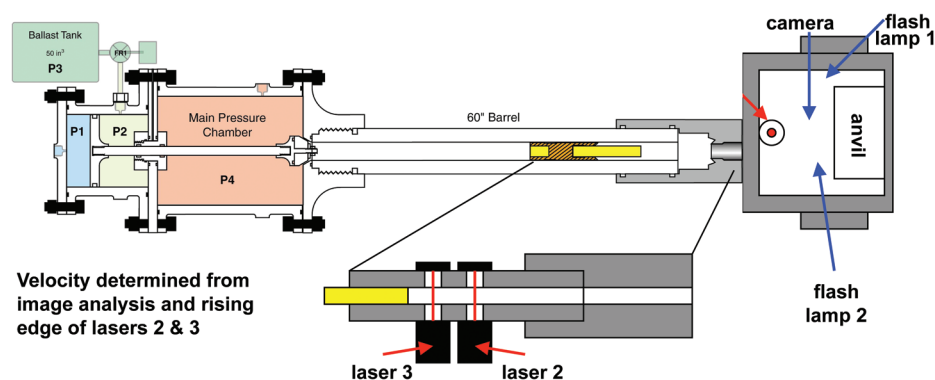
A schematic of the experimental setup is shown in Fig. 1. An oscilloscope tracks three lasers that are mounted on a composite-lined gun barrel and in the steel box containing the anvil to trigger diagnostics and determine the projectile velocity. A Phantom 7 camera allows images to be taken at 25 s per frame. We obtained ten shots on PBX 9501 and eight on PBXN-9 for impact speeds between approximately 80 and 214 m/s. In Fig. 2, two shots at nearly the same velocity show the quality of the information achieved in these experiments. A comparison of these images shows that a different behavior is observed in the more viscoelastic PBXN-9 than in the more brittle PBX 9501. For the same impact velocity, significantly more of the PBXN-9 is consumed in the fragmentation process than is the case with PBX 9501, while the PBX 9501 fragments tend to be much finer than in PBXN-9.

The quasi-static mechanical properties of PBX 9501 (Fig. 2 shows a representative plot) and PBXN-9 have been characterized in both tension and compression. In addition, LANL researchers [3] have used the Split Hopkinson Pressure Bar (SHPB) apparatus to measure the stress-strain curves at higher strain rates. This data (Fig. 3) is used to calibrate LANL's ViscoSCRAM explosives constitutive model. Together with ViscoSCRAM's shear crack-growth, a simple tensile failure condition that acts at the finite-element (FE) level has been added. The FE code ABAQUS is used to simulate the Taylor tests. The tensile failure condition used states that when an element is in hydrostatic tension and the principal tensile strain exceeds a specified value, the element will permanently cease to support a load. Using this technique, fragmentation was simulated. At early times (see Fig. 4) ViscoSCRAM

Fig. 1. Schematic of the Taylor impact setup used in this study.

Velocity determined from image analysis and rising edge of lasers 2 & 3

performed well, but at later times, the element death technique gave poor results, presumably attributable to the lack of fragment-fragment interactions. This work is to be viewed as a first step in modeling Taylor experiments.
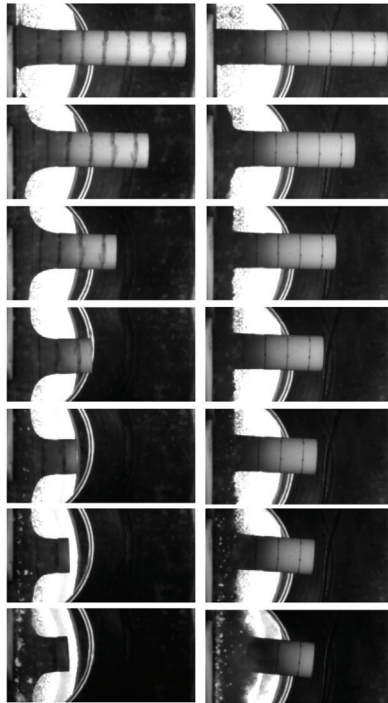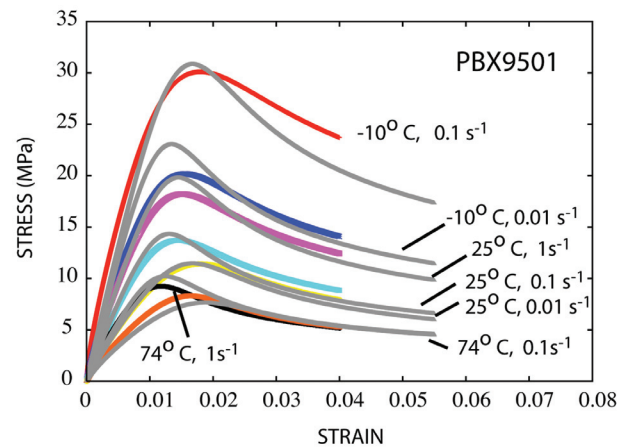

Fig. 2. PBXN-9 Taylor tests at a velocity of 136.6 m/s (left), and PBX 9501 at a velocity of 132.1 m/s (right). Each frame is 100 m/s in succession, with the first frame shown being at the instant of impact and the final frame being near the conclusion of the deformation.

Fig. 3. Low-rate uniaxial compression experiment and ViscoSCRAM model fits to PBX 9501 stress-strain curves. The theory curves are the longer ones.
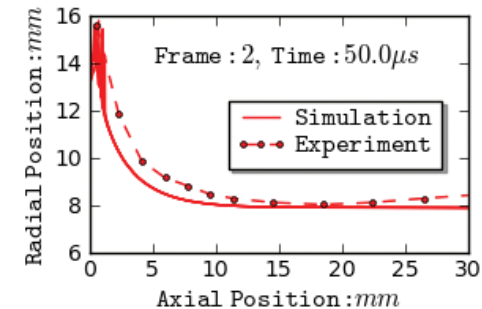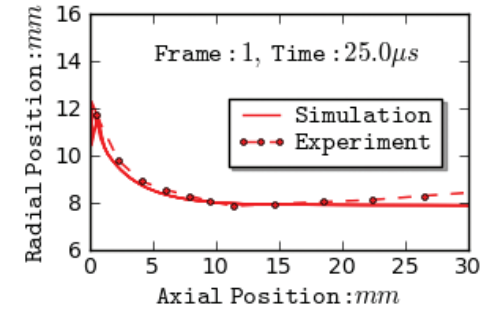
Fig. 4. PBXN-9 136 m/s experiment and simulated cylinder profiles at 25 and 50 ms. At 50 ms, the jagged simulation profile between 14 and 16 mm occurs due to fragmentation. A critical tensile strain of unity was used.

[1] Christopher, F.R. et al., *Proc 11th Int Detonation Symp* **286** (1998).
[2] Liu, C. and R. Ellis, LANL (unpublished).
[3] Gray, G.T. III et al., *Proc 11th Int Detonation Sym* **76** (1998).

Accelerated Test Methods for Reliability Prediction

David H. Collins, CCS-6; Jason K. Freels, Air Force Institute of Technology; Aparna V. Huzurbazar, CCS-6; Richard L. Warr, Air Force Institute of Technology; Brian P. Weaver, CCS-6

Traditional testing methods are costly and inefficient for predicting the lifetimes of materials that are expected to function reliably over many years, such as components in the nuclear stockpile. Alternatives to traditional tests accelerate the aging of components by subjecting them to conditions outside normal service or storage ranges, which, in principle, provides test data within a compressed time frame. Alternatives include accelerated life testing (ALT), accelerated degradation testing (ADT), highly accelerated life testing (HALT), highly accelerated stress screening (HASS), and variants on these methods.

Use of accelerated methods is often hindered by organizational conflicts between testing as part of an iterative process of finding and removing defects and testing as a means of estimating or predicting reliable life. We are developing a taxonomy for classifying types of accelerated tests and a statistical framework for reconciling conflicting objectives and optimizing the use of testing resources.

Many systems are subject to requirements for extremely high reliability over long periods of operation or storage—the nuclear stockpile is a prime example. In addition, systems and components often need to be developed within a time frame that is much shorter than their required reliable operating lives. These requirements present a challenge to traditional reliability engineering, in which items are tested to failure under expected operating conditions in order to predict the reliable lifetime of a deployed system. When lifetimes are measured in years or decades, this approach is no longer feasible.

Reliability can be defined as "the ability of an item to perform a required function, under given environmental and operational conditions and for a stated period of time" [1]. Formally, if $F(t)$ is the probability of failure at or before time t, the reliability function $R(t) = 1-F(t)$ gives the probability that the item will still be functional at time t. Derived quantities include the probability density of failure, $f(t) = dF(t)/dt$, and the hazard rate, $h(t) = f(t)/R(t)$. The hazard rate gives the failure rate at time t, given survival up to t—knowing whether $h(t)$ is increasing, decreasing, or constant is helpful in predicting an item's useful life.

A standard engineering rule of thumb is that the hazard rate for a complex system follows a "bathtub curve," as shown in Fig. 1. Early in the life of a system it may exhibit a high failure rate due to production defects or design flaws. It is expected that these problems will be corrected,

leading to a period where the failure rate is constant (and low). Near the end of its useful life, the failure rate is increasing due to "wearout," which might be wearing out of mechanical components, or some physical or chemical degradation process.

Reliability engineering has several goals, which can be characterized in terms of the bathtub curve. Ideally, the early part of the curve would occur before deployment of the system–this involves a planned process of reliability growth [2], illustrated in Fig. 2, where progress is measured by testing against a series of "learning curves" [3] as defects are identified and corrected, with major problems deferred to separate corrective action phases. This process requires predictive models for reliability growth and testing methods that rapidly identify defects—both of these areas are part of our ongoing research.

Testing during development should verify that the system can be deployed with an acceptably low failure rate. Assuming this can be done, an additional important requirement is to understand aging characteristics in order to predict the onset of wearout or degradation failures, either in operation or when the system is stored in a ready state. Figure 3 plots data from a degradation test, where metal alloy components are prepared with a small notch, then subjected to repeated flexing until fatigue cracks form and lengthen. The inset plot at the top is the estimated probability density of time (flex cycles) to reach a critical length where failure is imminent. This density can be used to predict the service life of the component before it must be replaced [4].

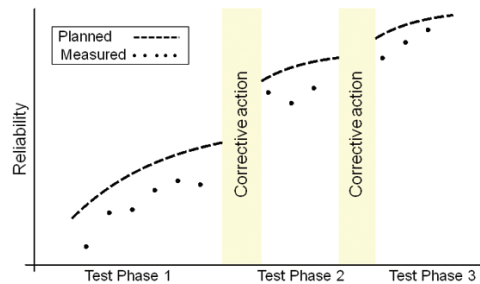*Fig. 1. "Bathtub curve" of lifetime hazard rate.*
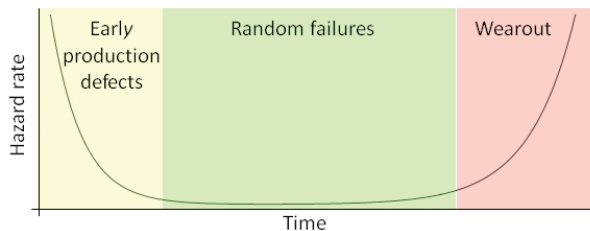
*Fig. 2. Reliability growth over a product development cycle.*

For highly reliable components and systems, testing of the sort just described cannot be completed within an acceptable time frame if it is performed under normal service conditions—collecting a sizable number of failures or sample degradation paths might take years. Approaches to avoiding this problem are accelerated life testing (ALT) and accelerated degradation testing (ADT), where items are tested under stress conditions of temperature, mechanical loading, vibration, power cycling, etc., that are outside the normal operating range, in order to produce degradation or failure in a shorter time [5]. ALT may include progression through a range of stresses, or application of multiple stressors simultaneously. At LANL, ALT/ADT has been applied to materials ranging from organic polymers [6] to plutonium [7] in order to predict how the aging of these materials will affect the components in which they are used.

The intent of ALT/ADT is to accelerate the production of failures or degradation that could occur in normal usage, so the range of stressors must not initiate novel failure modes–for example, a temperature high enough to melt solder on an electronic circuit board. In addition, for accelerated tests to have predictive value, the experimenter must be able to model how stress reduces the life of the item. As a typical example, the exponential distribution is often used to predict the life of electronic components: $R(t) = \exp(-\lambda t)$, where λ is the mean failure rate. The Arrhenius model for temperature dependence of reaction rates is used to predict changes in reliability when temperature is increased, by setting $\lambda = \lambda(\tau) = \alpha\exp(-\beta/\tau)$, where α and β are experimentally determined. Other models, such as the Eyring, can be used to account for the effect of multiple stressors.

Highly accelerated life testing (HALT), which uses combinations of stressors at progressively higher levels, while superficially similar to ALT, is aimed only at finding and correcting design faults during product development [8]. As practiced, it does not yield predictive statistical models [9]. Highly accelerated stress screening (HASS) is a related method applied to the elimination of defective items from a production process.

Fig. 3. Sample paths for fatigue crack growth.
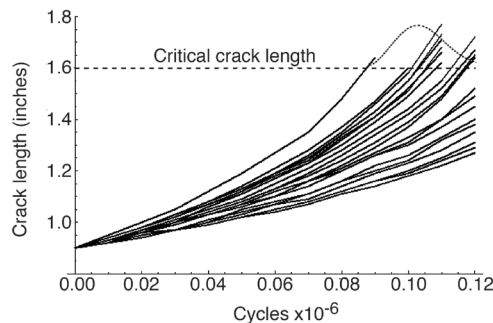
Our research suggests that although ALT and HALT testing involve similar experimental setups, and could in principle provide similar data, they are performed by different organizational units for different purposes. HALT has been shown to be effective for accelerating reliability growth, but contractual requirements for a demonstrated level of reliability may mandate statistically oriented methods such as ALT. This creates an organizational conflict and may result in suboptimal allocation of limited testing budgets.

We believe there are synergies waiting to be exploited between the various accelerated methods and are currently classifying types of accelerated tests and cataloging experimental design and analysis techniques. Based on these efforts, we hope to develop experimental protocols that provide results usable both by product engineers who are testing in order to find bugs in the design and by statisticians collecting data for lifetime prediction. In addition, we hope to develop robust models for accelerating factors that reduce the effort required to analyze ALT and HALT results. This effort will benefit the current B61 life extension program (LEP) as well as future LEPs and other long-term aging studies.

[1] ISO 8402, *Quality Vocabulary*, International Organization for Standardization, Geneva, Switzerland (1986).
[2] MIL-HDBK-189C, *Reliability Growth Management*, US Department of Defense, Washington, DC (2011).
[3] Duane, J.T., *IEEE Trans Aerosp* **2,** 563 (1964).
[4] Lu, C.J. and W.Q. Meeker, *Technometrics* **35,** 161 (1993).
[5] Nelson, W.B., *Accelerated Testing: Statistical Models, Test Plans, and Data Analysis,* John Wiley & Sons, Hoboken, NJ (1990).
[6] Labouriau, A., and T.S. Stephens, "Report on Assessment of Aging in Polymers," LANL Technical Report LA-UR-08-6800 (2008).
[7] Martz, J.C., and A.J. Schwartz, *J Miner Met Mater Soc* **55,** 19 (2003).
[8] Hobbs, G.K., *Accelerated Reliability Engineering: HALT and HASS,* John Wiley & Sons, Chichester, UK (2000).
[9] Nelson, W.B., *IEEE Trans Reliab* **54,** 194 (2005).

Modeling the Response of Single Crystals to High-strain-rate Deformation

Benjamin L. Hansen, Irene J. Beyerlein,
Curt A. Bronkhorst, T-3; Ellen K. Cerreta, MST-8;
Darcie Dennis-Koller, WX-9

The goal of this work is to formulate a constitutive model for the deformation of metallic single crystals over a wide range of strain rates, which is integral to computing reliable stress states of metallic polycrystals under shock loading. An elastic-viscoplastic, slip-based single-crystal model that accounts for crystallographic orientation, temperature, and strain-rate dependence has been formulated based on dislocation dynamics simulations and existing experimental data. Dislocation densities are separated according to type and given a systematic set of adaptable interaction rates. This allows the tracking of each dislocation event statistically. The direct use of dislocation density as a state variable also gives a measurable physical mechanism to strain hardening. The plastic model transitions from the low-rate, thermally activated regime to the high-rate, drag-dominated regime by use of a distribution of dislocation velocities, including kinetic effects. The model has been compared favorably with copper. The novelty of this model is that, for the first time, transition to drag-dominated dislocation motion is predicted rather than empirically fit to experimental data.

This document reports progress on a model for deformation of a single-crystal volume at a variety of strain rates from quasi-static to shock-loading. It includes temperature and pressure effects and the ability for improved validation information using dislocation densities. The end goal is to embed this single-crystal material model into a framework for simulating polycrystalline materials (such as finite element) in order to probe the variation in stress states caused by the microstructural differences. Information about these stress states leads to information about potential damage sites, which can be incorporated into the polycrystalline configuration to understand damage progression and improve modeling capabilities by incorporating salient features of the polycrystalline material.

Understanding the response of single crystals to high-strain rate deformation provides the pathway to predicting microstructural effects in shocked polycrystals (see Fig. 1). Such effects are required to determine grain boundary stress states, which are the dominant factor in void nucleation for pure materials.

While the dislocation interactions themselves are being modeled for the single-crystal in this work, for tractability it is desired to have a statistical average over the dislocation interactions. Resolution of the individual dislocations is computationally too intensive, but the average reactions that take place in a section of material may be modeled to provide a more physical representation of the deformation process. It is also noted that to continue to higher strain rates, the motion of dislocations is believed to become drag-limited rather than thermally activated, requiring a transition in the average dislocation mechanics with strain-rate changes.

While many models exist which use dislocations as a state variable, none appear to consider the distribution of dislocation velocities. The dislocation density work of Wang, Beyerlein, and Lesar [1] showed that a small portion of the dislocations moved at a high velocity, causing the majority of the plastic deformation, while most dislocations moved at a lower velocity. This motivated the idea of using a distribution of dislocation velocities to model plastic deformation.

The plastic deformation evolution is based on statistical dislocation populations and the evolution of those populations. The material point volume is selected to be sufficiently large so that dislocation populations can be adequately described by a line density. Three main dislocation populations are considered: 1) glissile dislocations, which are free to move (ρ^M); 2) glissile dislocations, which currently do not move due to being blocked by obstacles–these are referred to as "pile up" (ρ^P), and 3) sessile dislocations acting as debris in the material (ρ^D). Mobile

Fig. 1. Variation in single-crystal plastic behavior as evidenced by the average kernal misorientation in an electron back-scatter diffraction (ESBD) scan of copper that has been shocked for incipient damage. Scale is in degrees misorientation.

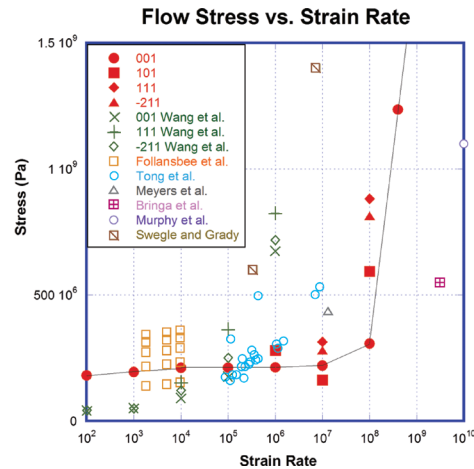**Flow Stress vs. Strain Rate**
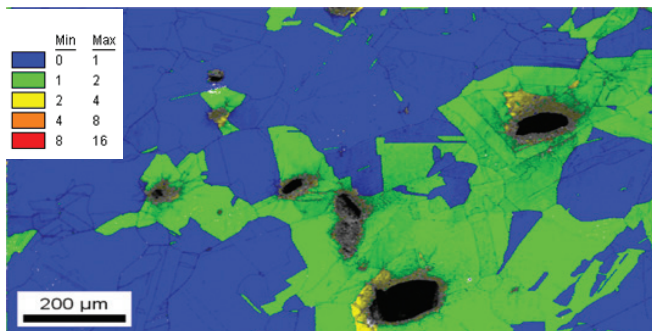
Fig. 2. Comparison of simulated flow stress at various strain rates to a variety of experimental Hopkinson bar, plate shock driven, laser driven, molecular dynamics, and dislocation dynamics simulations. Data taken from Wang, Beyerlein, and Lesar [2] and Armstrong and Zerilli [3].

dislocations (ρ^M) move through the crystal and are either blocked by obstacles, react and annihilate, react and become sessile, or exit the region. The population $\rho\{^P\}$ is distinguished as glissile, but is temporarily rendered immobile because it is stopped by other dislocations, substructure, or other impediments to dislocation glide, such as grain boundaries or inclusions. Pile-up dislocations can become mobile, react and annihilate, or react and become sessile. The population of dislocations ρ^D contains dislocations that are sessile and do not contribute to plastic strain–they are permanently immobile and are restricted to short thermal migratory motion. These sessile dislocations can be statistically stored or in an organized substructure. Statistically stored dislocations are randomly distributed. Substructure dislocations are concentrated in planar volumes.

The separation of dislocations into the mobile and pile-up populations is motivated by the dislocation dynamics work of Wang, Beyerlein, and LeSar, which concluded that a few elite dislocations carry the bulk of plastic deformation, where most dislocations have a comparatively low velocity. The motivation to model the distribution of dislocation velocities is used to create the simplest distribution of two populations moving at two velocities. Again for simplicity, one of the velocities is assumed to be insignificant and is approximated as zero. This differs from most current dislocation-based rate-sensitive plasticity models, which assume the entire dislocation population moves at the same velocity.

With parameters set to physical order of magnitude, the results of single-crystal tensile tests at various rates are compared to experimental copper tests. Results are seen in Fig. 2 for a single strain rate showing the prediction of the uniaxial stress state and dislocation density. Figure 3 shows that simulations predict an appropriate transition from thermally activated to drag-dominated dislocation motion, given the level of uncertainty in experimental data.

The key findings of this work to date are: 1) expansion of current direct simulation of polycrystals to include a single-crystal model capable of transitioning from low to high strain rates, and 2) prediction of the transition from thermally activated to drag-dominated strain rates. Further investigation is easily adaptable in the model. The model, when coupled with virtual polycrystal representations, allows for prediction of grain-boundary stress during shock experiments for determination of critical microstructure characteristics for damage.

001 Single Crystal Copper at rate 1E6
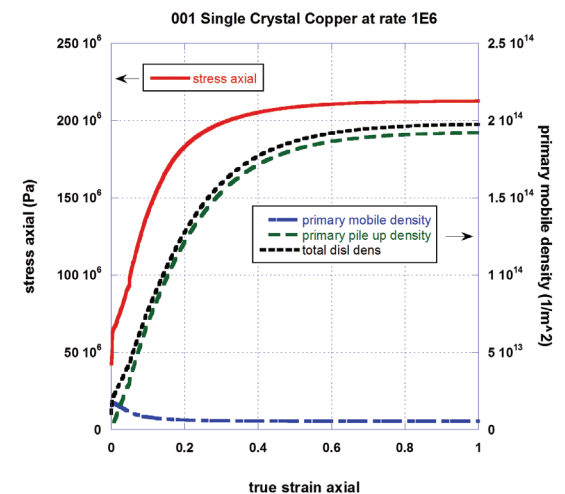
Fig. 3. Simulations of single crystals under tensile stresses using physically realistic parameters for copper were conducted to implement and validate the single-crystal model.

[1] Wang, Z.Q. et al., *Phil Mag* **87,** 2263 (2007).
[2] Wang, Z.Q. et al., *Int J Plast* **25,** 26 (2009).
[3] Armstrong, R.W. and F.J. Zerilli, *J Phys Appl Phys* **43,** 492002 (2010).

Developing More Efficient Fuel Cells Using Computational Methods

Neil J. Henson, Ivana Matanović, T-1;
Fernando H. Garzon, MPA-11; Paul R. Kent, ORNL

In this project we use a number of state-of-the-art, highly parallelized computational-chemistry tools in order to rationalize the fundamental chemical processes occurring at the atomic level in a fuel cell. This knowledge will enable the optimization of materials properties to develop more efficient and robust fuel-cell designs. First-principles quantum-chemistry techniques and more approximate molecular-mechanics-based methods using empirical potentials are employed in parallel with experimental measurements for validation.

We carried out an extensive computational study of the structure, reactivity, and stability of three different platinum-nickel (Pt-Ni) alloys, Pt_3Ni, PtNi, and $PtNi_3$ using the Vienna Ab Initio Simulation Package (VASP) [1], a periodic density functional theory code. Our aim was to determine the effect of the subsurface layer composition on the catalytic activity of the platinum surface. These alloys are candidates for the replacement of pure platinum electrodes in fuel cells used to convert hydrogen and oxygen fuel to water and electricity [2]. Since pure platinum electrodes are costly, one method for alleviating this cost is to alloy platinum with base metals, such as nickel. The research question is how this alloying process will impact activity–this computational study attempts to answer this question.

We constructed models for the alloy materials that featured surfaces to which the molecules taking part in the chemical reactions could bind. The surface models featured three types of slabs of platinum and nickel containing a platinum surface monolayer supported on a second layer containing 50%, 100%, and 75% of nickel, respectively, with a bulk layer below to represent alloys with different concentrations of platinum.

Equilibrium adsorption potentials represent the strength of binding of the molecules to the surface. We calculated these potentials for the oxygen reduction reaction (ORR) intermediates–this is the fundamental chemical step in the fuel-cell chemistry that controls the reaction rate and, therefore, the efficiency of the fuel cell process at the surface. The results of these calculations provide a method for constructing free-energy diagrams for the ORR mechanism and for gauging catalytic activity (Fig. 1).

We also addressed the critical question of the stability of these materials in an aqueous environment. This provides an indication of the lifetime of the fuel-cell materials in a working device. The calculations identify the most stable state of the surface as a function of pH and potential, which reflects the working conditions of a fuel cell. The (111) surface of all three models of Pt-Ni alloys exhibits improved oxygen reduction activity compared with that of pure Pt (111). Our calculations show that the ORR over-potential, which is a measure of the fuel-cell efficiency, decreases in the order Pt (0.55 V) > Pt_3Ni (0.24 V) > $PtNi_3$ (0.19 V) > PtNi (0.15 V) [3]. We therefore conclude that the catalytic activity for ORR will increase as Pt < Pt_3Ni < $PtNi_3$ < PtNi, and find that the largest improvement occurs for a PtNi alloy with 100% nickel in the second layer. We also predict that PtNi is the least susceptible to corrosion at similar pH and cell potentials based on the calculated shifts of the electrochemical dissolution potentials for the Pt-Ni alloys relative to platinum with values of −0.27 V for $PtNi_3$, +0.13 V for Pt_3Ni, and +0.30 V for PtNi [3].

In addition to this work on bulk metallic structures, we are also investigating the structure and properties of platinum nanotubes as another method of reducing platinum content [4]. We have calculated the relative stability of a large number of platinum (n,m) nanotubes ranging in diameter from 0.3 to 2.0 nm in the gas phase and water environment (Fig. 2). We have also calculated the adsorption energies of intermediates in the ORR (oxygen and hydroxyl) and used them to

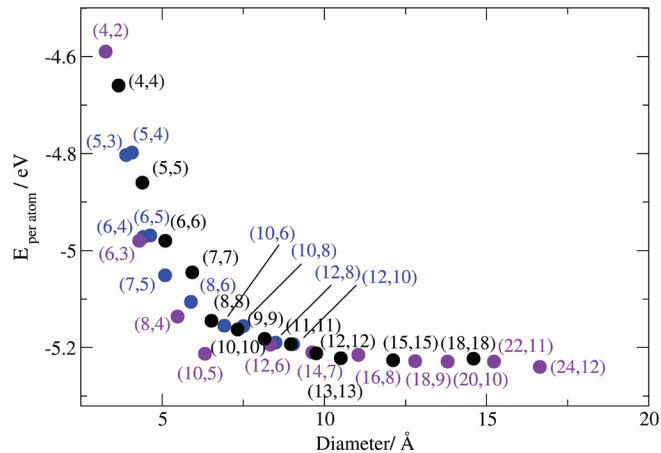Fig. 1. Oxygen reduction reaction on Pt-Ni alloys.
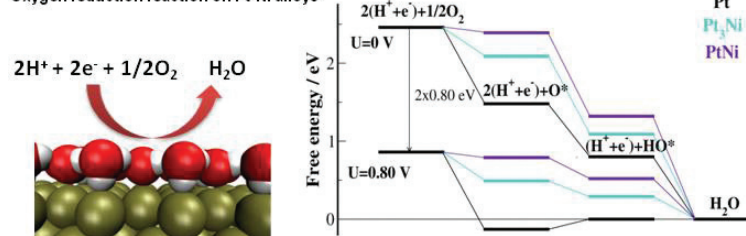
Fig. 2. Calculated stability of platinum nanotubes as a function of diameter.

construct free-energy diagrams for the oxygen reduction dissociation mechanism. Based on these calculations we conclude that the smaller Pt nanotubes (~0.5 nm in diameter) have a huge over-potential for ORR and thus prove to be very poor catalysts for ORR. However, bigger Pt tubes (>1 nm in diameter) have a lower over-potential than bulk platinum for up to 150 meV, indicating that these materials might be better ORR catalysts than bulk Pt.

The stability of the platinum nanotubes in the gas phase is investigated in terms of cell potentials at which the tube will start to dissolve electrochemically. The most stable state of the nanotubes as a function of pH and potential is addressed by calculating its Pourbaix diagrams (Fig. 3). Tubes with n=m chirality and diameter >1 nm have been shown to endure the highest cell potentials; however, these are still for a 130-meV lower than the cell potentials at which bulk platinum will start to dissolve. In conclusion, we predict that some platinum nanotubes show improved catalytic activity compared to bulk platinum, but they may be more susceptible to corrosion.
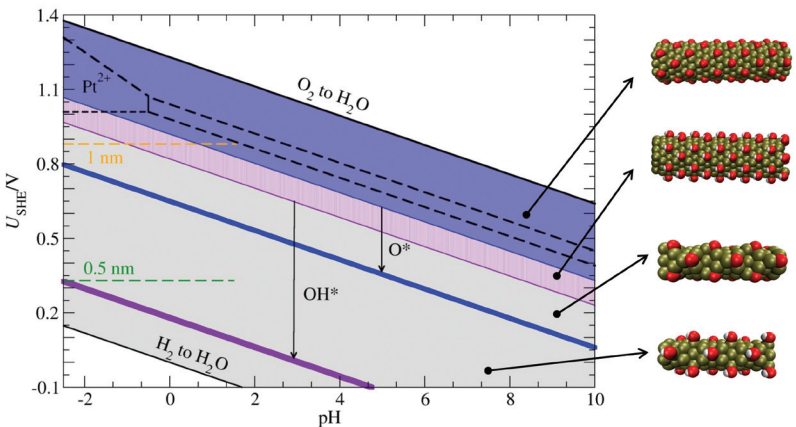
Fig. 3. Pourbaix diagram of platinum nanotubes compared to the bulk (black dashed lines). The regions of oxygen and hydroxide surface adsorptions are shown in blue and violet. The green (orange) dashed lines show the solubility boundary for Pt nanotubes with diameters of 0.5 and 1 nm.

[1] Kresse, G. J. and Hafner, *J Phys Rev B* **47,** 558 (1993).
[2] Gewirth, A. and M.S. Thorum, *Inorg Chem* **49,** 3557 (2010).
[3] Matanovic, I. et al., *J Phys Chem C* **115,** 10640 (2011).
[4] Oshima, Y. et al., *Phys Rev B* **65,** 121401 (2002).

Nonlocal Continuum Modeling of Crystalline Plasticity

Jason R. Mayeur, Hashem M. Mourad,
Irene J. Beyerlein, T-3

Ongoing efforts devoted to the development of physics-based nonlocal models of crystalline plasticity are discussed with respect to modeling the mechanical response of copper-niobium (Cu-Nb) lamellar nanocomposites. Two different modeling strategies are put forward, and results demonstrating the scale-dependent mechanical behavior of an idealized polycrystalline material are presented. The results indicate a "smaller is stronger" trend and offer promise that these methods can be successfully adapted to model the Cu-Nb lamellar composite material system.

The limitations of classical (local) continuum modeling approaches under certain conditions have been established for quite some time now. For example, there are pathological mesh sensitivity issues that arise when material softening and/or geometric constraints lead to strain localization and models are unable to predict the size-dependent mechanical behavior that is often observed in submicron-sized material systems. These shortcomings of local continuum models are the direct consequence of the scale invariance of the governing equations, that is, fundamental length-scale parameters are absent. Several methods of incorporating nonlocality into continuum models have been proposed and pursued to various extents in an effort to address these limitations. Nonlocal continuum modeling approaches may be broadly classified into three major categories: 1) integral formulations (strongly nonlocal), 2) gradient formulations (weakly nonlocal), and 3) multiphysics formulations. In the present work, we are developing both gradient and multiphysics models for crystalline plasticity.

The primary application of interest for our model development efforts is the simulation of the synthesis and mechanical response of bimetallic lamellar nanocomposites. The initial material system being studied is a copper-niobium (Cu-Nb) composite that has been a material of interest at LANL for the past 10 to 15 years [1]. The earlier studies were conducted on composites that were synthesized using physical vapor deposition (PVD), whereas current efforts are focused on samples that are produced using an accumulative roll-bonding (ARB) technique [2]. By employing the ARB manufacturing process, larger sample sizes can be produced and layer thicknesses ranging from 10 nm to several microns can be achieved with relative ease, thereby enabling the study of the overall deformation response of the composite as a function of layer thickness. As the layer thickness is refined below the submicron level, this material system exhibits elevated strength and radiation damage resistance as compared to composites with larger layer thicknesses. We hypothesize that the transition in deformation modes is inherently related to the Cu-Nb interfaces and their increased participation in the overall inelastic deformation of the composite as the layer thickness is reduced. For larger layer thicknesses, dislocation motion in the bulk dominates the total material response; however, as the layer thickness is refined, the ability of the interfaces to absorb, transmit, and block dislocations plays the dominate role. The underlying physics of dislocation interactions at the bimaterial interfaces, while different, is analogous to what transpires at grain boundaries in polycrystalline metals. The most physical and accurate way to model these processes is through the development of a multiphysics type of nonlocal model that couples standard continuum crystal plasticity theory with an additional conservation equation governing dislocation transport. The length scales that enter this model formulation are directly related to the dislocation transport. The development of the coupled plasticity/dislocation transport model capable of describing the complex dislocation-interface interactions in the Cu-Nb lamellar composites is a work-in-progress.

Ongoing research is also working to extend current nonlocal modeling capabilities by adapting them to the application of interest. In this vein, a nonlocal crystal plasticity model that treats lattice curvature (a second gradient of deformation) in addition to the lattice strain as a fundamental deformation measure and constitutive response variable is being tailored to model the Cu-Nb composite material system. This type of weakly nonlocal model is a more coarse-grained description as compared to the dislocation transport model and is more computationally efficient,
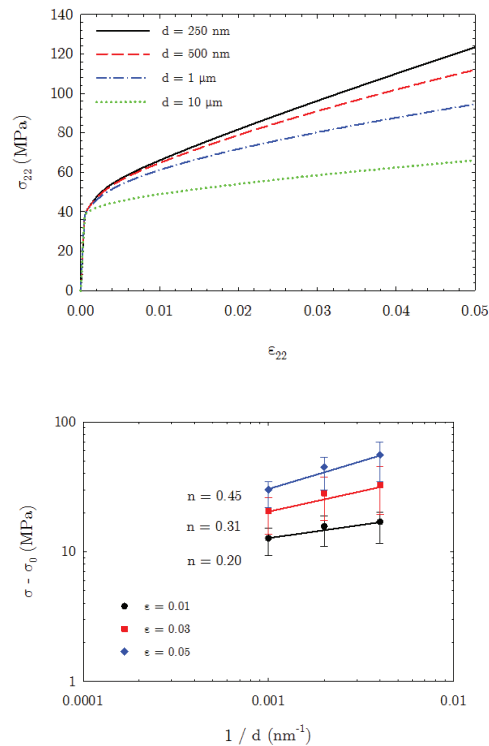
Fig. 1. (a) Nominal stress-stress curves for an SVE. (b) Scaling of nominal stress as a function of grain size for various levels of applied strain.

but makes some sacrifices with regard to its true predictive capabilities. The importance of accounting for the effects of lattice curvature and geometrically necessary dislocations (GND) on the mechanical response of small scale material systems has long been established [3], and there are many different approaches for capturing these effects. The grain size-dependence of the stress-strain response of an idealized polycrystalline material is investigated to demonstrate the capabilities of our approach. Small statistical volume elements (SVE) containing 30 randomly oriented grains are subjected to remote uniaxial loading under plane strain conditions. Periodic boundary conditions have been employed within the plane. The grains are modeled as regular hexagons and four different grain sizes are considered, d = 250 nm, 500 nm, 1 mm, and 10 mm. The nominal stress-strain curves for the four different grain sizes are given in Fig. 1a for one of the SVEs considered; Fig. 1b shows the scaling of the flow stress with the inverse grain size at different levels of applied far-field strain. The expected "smaller is stronger" type of behavior is predicted by the model, and the scaling exponents of the flow stress are observed to increase with deformation. The scaling exponents range from n = 0.2–0.45 for the range of applied strains considered, and are within reasonable proximity of the traditional Hall-Petch flow stress scaling exponent of n = 0.5. In Fig. 2, the magnitude of the GND density is plotted for two different grain sizes and we observe that the intensity of the GND fields increase with diminishing grain size and are concentrated within the grain boundaries. This is exactly the response characteristics that would be expected based on prior experimental and analytical results, and also on the fact that lattice curvature (GNDs) is most pronounced at grain boundaries. These results capture the correct trends in scale-dependent mechanical response and offer promise that these methods can be successfully adapted to model the Cu-Nb lamellar composites.
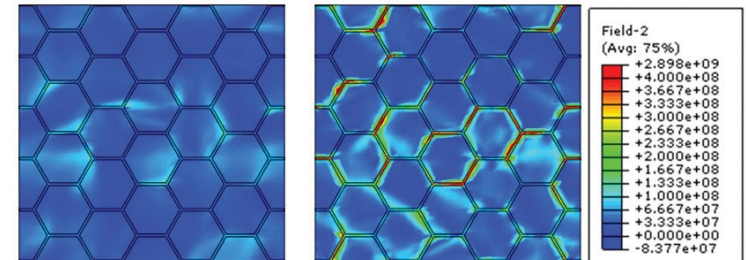
Fig. 2. Contours of geometrically necessary dislocation density in units of mm-2 (left) d = 10 mm (right) d = 250 nm.

[1] Misra, A. et al., *Scripta Mater* **39,** 555 (1998).
[2] Mara, N. et al., *Scripta Mater* **50,** 803 (2004).
[3] Ashby, M.F., *Phil Mag 21,* **399** (1970).

Finite-Element Formulation for Dynamic Strain Localization and Damage Evolution in Metals

Hashem M. Mourad, Curt A. Bronkhorst,
Thomas R. Canfield, Francis L. Addessio, T-3

Due to their inability to represent strain localization correctly, conventional computational methods yield mesh-dependent results in ductile failure problems. In the computational framework presented here, localization behavior is predicted using a material stability analysis and localized deformation modes are represented accurately and efficiently by embedding localization bands within larger computational elements. This framework also allows the use of different constitutive models inside and outside the localization band.

The formation of strain localization bands is often observed in metals undergoing high-rate plastic deformation. The material in these narrow bands undergoes intense plastic straining. In many cases strain localization is caused by–and subsequently interacts with–material softening mechanisms (e.g., void nucleation and growth, thermal softening), leading ultimately to failure. Thus, the effective treatment of ductile failure problems in a computational setting requires that localized deformation modes be represented accurately.

With conventional finite element techniques, this can only be achieved by resolving localization bands explicitly, that is, via mesh refinement, which is prohibitively expensive given the small width of a typical localization band (10–30 μm) compared to the characteristic dimension of a specimen (~10 mm) or structure (~1 m). In addition, conventional methods are known to yield mesh-dependent results, for example, predicted localization bands tend to follow mesh lines.

To circumvent these difficulties, an explicit finite-element formulation for dynamic strain localization was developed, based on the assumed-strain technique of Belytschko et al. [1]. In this formulation, a material stability analysis is used to check for incipient localization behavior at each material (Gauss quadrature) point. Notably, the location and orientation of the nascent localization band are also determined by the same stability analysis, instead of being dictated by the dimensions or orientation of the mesh.

In the absence of localization, the strain field within a given element remains continuous. Once the onset of localization is detected, however, the discontinuous strain mode associated with the nascent localization

band is added to the element's strain field and is allowed to evolve gradually. This approach ensures a smooth transition from uniform to localized deformation. In addition, embedding localization bands in this manner, within significantly larger computational elements, allows spatially converged solutions to be obtained at a reasonable computational cost. It is also important to note that the width of the localization band is treated as a material parameter, completely independent of the mesh size.

As a numerical example, we consider a rectangular block (length = 0.14 m, width = 0.1 m) subjected to plane strain extension. The material is assumed to follow the elasto-plastic material law illustrated in Fig. 1, and has the following material properties: Young's modulus E = 200 GPa, Poisson's ratio v = 0.29, isotropic hardening/softening moduli Kh = 220 MPa, Ks = -66 MPa (with a negative value signifying softening), yield strength σ_y = 300 MPa, ultimate strength σ_u = 310 MPa, and mass density ρ = 7850 kg/m^3.

Exploiting symmetry to reduce computational cost, only the upper right quadrant of the block is modeled. An upward velocity of 2 m/s is applied to the upper boundary, as shown in Fig. 2. The localization band is assumed to have width b = 3.33 mm. The material strength is reduced in the bottom-left element (σ_y = 200 MPa, σ_u = 210 MPa) to simulate the existence of a material imperfection, making this element a favorable nucleation site for localization bands.

To examine the ability of the formulation in alleviating mesh dependency, the problem is first solved using a coarse mesh consisting of 5×7 elements, and then repeated using a mesh including 10×14 elements.

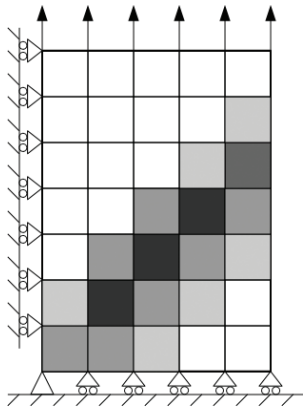Fig. 1. Illustrated elasto-plastic material law.

The results of these two simulations are then compared to the reference solution that is obtained by explicitly resolving the localization band using a fine mesh consisting of 15×21 conventional elements.

The load-deflection curve and the deformed shape of the block, as predicted by each of the three meshes, are shown in Figs. 3 and 4, respectively. It is clear from Fig. 3 that the formulation is successful in capturing the correct response of the overall structure throughout the loading process, including the softening regime. It can also be seen from Fig. 4 that the location and orientation of the localization band, and its effect on the overall deformation of the structure, are captured reasonably well, even by the coarsest mesh.

Fig. 2. Exploiting symmetry to reduce computational cost, only the upper right quadrant of the block is modeled. An upward velocity of 2 m/s is applied to the upper boundary. Shaded area represents the localization band that forms as the block deforms. Elements where the material stability analysis indicates that localization behavior was initiated at some stage of the loading process, at all (four) Gauss points, are shown in the darkest shade. Lighter shades signify instability at a smaller number of Gauss points.

Step: Step-1
Increment 400568: Step Time= 4.0000E-03

Deformed Var: U   Deforamtion Scale Factor: +1.000e+00

Fig. 4. The deformed shape of the block, as predicted by each of the three meshes.

Fig. 3. The load-deflection curve, as predicted by each of the three meshes.

[1] Belytschko, T. et al., *Comput Meth Appl Mech Eng* **70,** 59 (1988).

Creep in Zirconium Described with Fully Atomistic Simulations Coupled to VPSC Models

Gopinath Subramanian, Danny Perez, T-1;
Blas P. Uberuaga, Carlos N. Tome, MST-8;
Arthur F. Voter, T-1

Zirconium is the most widely used element for cladding material in pressurized water reactors. The cladding is the last confinement barrier that prevents unstable atom release and is susceptible to creep under the influence of internal pressure and temperature due to residual fuel power and radiation. Atomistic simulations are used to observe phenomena, such as diffusion and aggregation of interstitials and vacancies, and parameterize the Visco Plastic Self Consisten (VPSC) model, which describes creep at experimentally relevant length and time scales.

Creep is a time-dependent irreversible deformation that occurs at a low rate and at stresses below the ultimate stress of the material. Non-hydrostatic stresses (i.e., shear stresses) are required for creep. Two different kinds of creep are relevant in cladding materials, namely thermal creep and irradiation creep. Both these deformation processes require elevated temperatures and applied stresses, while the latter also requires the presence of irradiation.

The Visco-Plastic Self Consistent (VPSC) model [1] describes creep on the grain level (micron scale), by treating each grain as if it is immersed in an effective medium. This model is attractive because of its predictive capability over experimentally relevant time and length scales. It can provide useful design guidelines, direct experiments, and aid in the phenomenological interpretation of measurements. However, parameterizing VPSC requires a large amount of non-trivial information about the system. Traditionally, either of two approaches has been employed: 1) greatly simplified models of physically relevant systems are used to extract these parameters, or 2) parameters are fit to

experimental data. As a result, descriptions of creep are usually valid only in a limited range of conditions, with severely restricted predictive capability.

In the present work, we use state-of-the-art accelerated molecular dynamics (AMD) simulation methodologies developed at LANL to significantly extend the time scales accessible by direct fully-atomistic simulations. These lead up to a complete parameterization of a VPSC description of the system that contains no adjustable parameters. This enables us to assess the prevalence of various mechanisms that contribute to creep, and the accuracy of different theories. We chose to study zirconium (Zr) in particular because of its relevance to LANL missions and its industrial prevalence.

VPSC models of creep in Zr require the rate at which point defects (interstitials, vacancies) are absorbed by line defects (dislocations, dislocation loops, and vacancy loops). We use existing Zr potentials [2], developed using Density Functional Theory (DFT), to characterize the dynamical processes that lead to defect absorption by line defects. This involves mapping out various interconversion and diffusion processes that point defects can undergo, and the energy barriers associated with them, as these determine the diffusive behavior of defects. The effect of strains (either external, or due to the presence of line defects) on these interconversion rates is incorporated by computing dipole tensors. These dipole tensors are a characteristic property of a defect and describe the energy gained by the defect in the presence of a strain field. Figure 2 shows the energy landscape of an interstitial in the strain field of an edge dislocation. We use molecular statics to compute dipole tensors, and molecular dynamics (MD) and AMD to discover

Fig. 1. Schematic of the principle behind VPSC. The deformation of a single grain is computed by replacing the surrounding grains by an effective medium. This process is repeated for all grains in the system.
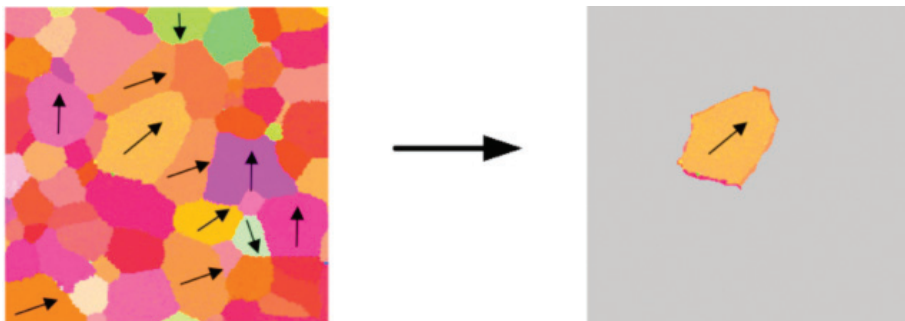
and characterize interconversion processes. In particular, we use Temperature Accelerated Dynamics (TAD) [3]—a specialized method that is built upon Harmonic Transition State Theory (HTST). TAD involves running MD at elevated temperatures, and extrapolating the results to lower temperatures. This results in the system exploring the potential energy surface orders of magnitude faster than it would have otherwise done, without loss of atomistic detail. We also characterize defect interconversion rates by computing the energy barriers and prefactors associated with each transition [4]. The catalog of events and barriers thus obtained will be fed into a Kinetic Monte Carlo simulation, to compute the net absorption rate of point defects by a variety of line defects. These rates will be used to compute dislocation climb rates, which form part of the parameterization of the VPSC description of creep.
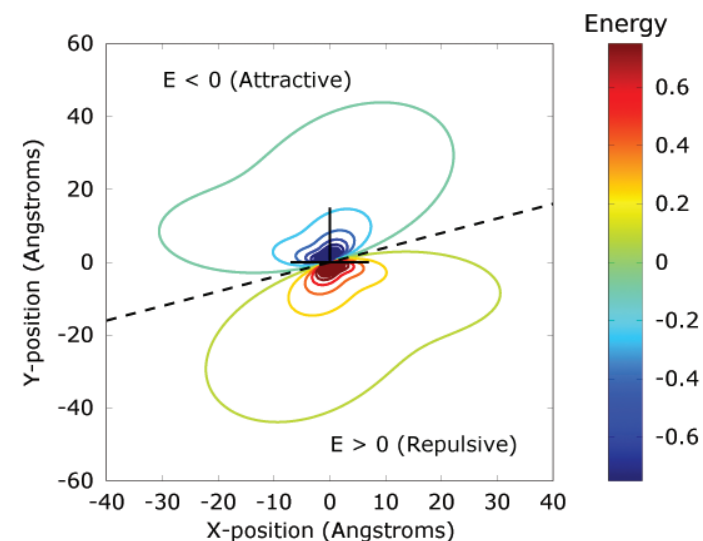
Fig. 2. Energy landscape of an interstitial due to the strain field around an edge dislocation. One side of the dislocation attracts interstitials, while the other repels them.

[1] Willaime, F., *J Nucl Mater* **323,** 205 (2005).
[2] Lebensohn, R.A. and C.N. Tomé, *Acta Metall Mater* **41,** 2611 (1993).
[3] Sørensen, M.R. and A.F. Voter, *J Chem Phys* **112,** 9599 (2000).
[4] Vineyard, G.H., *J Phys Chem Solid* **3,** 121 (1957).

Author Cross Reference

A

B

C

D

E

F

G

Q

R

S

T

U

V

W

X

Z

Organizational Abbreviations (for this publication only)

ADTSC
Associate Directorate Theory, Simulation, and Computation

ACS-PO
Advanced Computing Solutions Program Office

Applied Computational Physics Division
XCP-1 Lagrangian Codes
XCP-5 Materials and Physical Data

Applied Theoretical Design Division
XTD-6 Thermonuclear Applications Physics

Chemistry Division
C-PCS Physical Chemistry and Applied Spectroscopy

Decision Applications Division
D-DO Decision Applications Division Office

Computer, Computational, and Statistical Sciences Division
CCS-2 Computational Physics and Methods
CCS-3 Information Sciences
CCS-6 Statistical Sciences
CCS-7 Applied Computer Science

High Performance Computing Division
HPC-1 Scientific Software Engineering
HPC-3 High Performance Computer Systems
HPC-5 System Integration
HPC-DO High Performance Computing Division Office

LANL Institutes
INST-OFF LANL Institutes Office

Materials Physics and Applications Division
MPA-11 Sensors and Electrochemical Devices

Materials Science and Technology Division
MST-8 Structure/Property Relations

Nuclear Nonproliferation Division
N-2 Advanced Nuclear Technology

PADSTE
Principal Associate Directorate for Science, Technology, and Engineering

Theoretical Division
CNLS Center for Nonlinear Studies
T-1 Physics and Chemistry of Materials
T-2 Nuclear and Particle Physics, Astrophysics and Cosmology
T-3 Fluid Dyamics and Solid Mechanics
T-4 Physics of Condensed Matter and Complex Systems
T-5 Applied Mathematics and Plasma Physics
T-6 Theoretical Biology and Biophysics
T-DO Theoretical Division Office

Weapons Experiments Division
WX-7 High Explosives Science and Technology
WX-9 Shock and Detonation Physics

DoD Department of Defense
DOE Department of Energy
NNSA National Nuclear Security Administration
ANL Argonne National Laboratory
INL Idaho National Laboratory
LLNL Lawrence Livermore National Laboratory
ORNL Oak Ridge National Laboratory
PNNL Pacific Northwest National Laboratory
SNL Sandia National Laboratories

NCAR National Center for Atmospheric Research
NETL National Energy Technology Laboratory
NIF National Ignition Facility
NIST National Institute of Standards and Technology
NOAA National Oceanic and Atmospheric Administration
NRC Nuclear Regulatory Commission

Term	Definition	Term	Definition
1D, 2D, 3D, etc.	One-dimensional, two-dimensional, three-dimensional, etc.	DDT	Deflagration-to-detonation
A	Adenosine	DE	Differential equation
ADT	Accelerated degradation testing	DFT	Density functional theory
ALE	Arbitrary Lagrangian Eulerian	DNA	Deoxyribonucleic acid
ALT	Accelerated life testing	DNS	Direct numerical simulation
AMD	Accelerated molecular dynamics	DRAM	Dynamic random-access memory
AMO	Atlantic multidecadal oscillation	EEDF	Electron energy distribution function
AMOC	Atlantic meridional overturning circulation	EOS	Equation of state
AMR	Adaptive mesh refinement	ESBD	Electron back-scatter diffraction
API	Application programming interface	ESC	Enhanced surveillance campaign
Ar	Argon	FaST	Faculty and Student Team Program
ARB	Accumulated roll-bonding	FD	Finite difference
ASC	Advanced Simulation and Computing Program	FE	Finite Element
BB	Blackbody	Fe	Iron
bcc	Body-centered-cubic	FEL	Free electron laser
BDE	Bond dissociation energies	FEM	Finite element method
BH	Black hole	Fl	Fluorine
C	Carbon, or	FPI	Fixed point iteration
	Cytosine	FSI	Fluid-structure interaction
CASL	Consortium for Advanced Simulation of Light Water Reactors	G	Guanine
CCS	Carbon capture and storage	GB	Grain boundary
CCSI	Carbon Capture Simulation Initiative	GDP	Guanosine triphosphate
CFQ	Default Linux disk scheduler	GMRES	Generalized minimum residual method
Chandra	X-ray observatory	GND	Geometrically necessary dislocation
CO_2	Carbon dioxide	GPS	Global positioning system
COSIM	Climate, Ocean and Sea Ice Modeling project	GPU	Graphics processing unit
COTS	Commercial off-the-shelf	GRB	Gamma-ray burst
CPU	Central processing unit	Grb2	Growth receptor bound protein-2
Cr	Chromium	GrIS	Greenland ice sheet
CRADA	Cooperative research and development agreement	GT	Genital tract
Cu	Copper	GTRF	Grid-to-rod fretting
CUDA	NVIDIA's parallel computing architecture	H	Hydrogen
DAX-1	dosage-sensitive sex reversal	HALT	Highly accelerated life testing
DC SQUID	Direct current superconducting quantum interference device		

Term	Definition	Term	Definition
HASS	Highly accelerated stress screening	MR	Measurement resonator
HB	Hydrogen bond	MSA	Microstrip SQUID amplifier
He	Helium	N	Nitrogen
HE	High explosives	NADW	North Atlantic deep water
HIV-1	Human immunodeficiency virus 1	NaI	Sodium iodide
HMX	Octagen	Nb	Niobium
HO	High order	NDI	Nuclear data interface
HTST	Harmonic transition state theory	Ne	Neon
HW	Hardware	NEAMS	Nuclear Energy Advanced Modeling and Simulation Program
I/O	Input/output	Ni	Nickel
ICF	Inertial confinement fusion	NIRCam	Near-infrared camera
IFBA	Integral fuel burnable absorber	NIRSpec	Near-infrared spectrograph
IOP	I/O operations per second	NKA	Nonlinear Krylov acceleration
IR	Infrared	NLTE	nonlocal thermodynamic equilibrium
JFNK	Jacobian-Free Newton-Krylov	NORM	Naturally occurring radioactive material
JJ	Josephson junction	O	Oxygen
JWST	James Webb Space Telescope	OI	Optimum interpolation
k-eff	k-eigenvalue	ORR	Oxygen reduction reaction
Kr	Krypton	OUQ	Optimal uncertainty quantification
LBA	Logical block addresses	Pan-STARRS	Panoramic Survey Telescope and Rapid Response System
LEP	Life extension program	PBX	Plastic-bonded explosives
LEP	Life extension program	PCM	Phase-change memory
LHS	Latin hypercube sample	PCS	Predictor-corrector scheme
lncRNA	Long noncoding RNA	PD	Parkinson's disease
LO	Low order	PDE	Partial differential equations
LSST	Large Synoptic Survey Telescope	PDF	Probability density function
LTE	Local thermodynamic equilibrium	PEI	Polyethyleneimine
LWR	Light water reactor	PFC	Plasma-facing component
MD	Molecular dynamics	pH	Measure of acidity or basicity of a aqueous solution
MD	Molecular dynamics	PI	Power iteration
MFD	Mimetic finite difference	PLFS	Parallel log-structure file system
MHD	Magnetohydrodynamics	PMI	Predictive maturity index
MIRI	Mid-infrared instrument	Pop III	Primordial
		PPPM	Particle-particle-particle-mesh

Term	Definition	Term	Definition
Pt	Platinum	SW	Software
PVFS2	A parallel virtual file system	T	Thymine
PWR	Pressurized water reactor	TAD	Temperature accelerated dynamics
QA	Quality assurance	TFIID	Transcription factor
QL	Qubit loop	TGA	Thermogravimetric analysis
QMU	Quantification of margins and uncertainties	Ti	Titanium
QoS	Quality of service	TMT	Thirty-meter telescope
RMSE	Root-mean-square error	TRS	Trinucleotide repeat sequence
RTI	Rayleigh-Taylor instability	TRT	Thermal radiative transfer
RZ	Geometry type	U	Uranium
S	Sulfur	UO_2	Uranium dioxide
SBO	Shock breakout	UQ	Uncertainty Quantification
SCR	Scalable checkpoint/restart	UTA	Unresolved transition array
SCVT	Spherical centroidal Voronoi tessellations	UV	Ultraviolet
SDS	Sodium dodecyl sulfate	UVOIR	Ultraviolet/optical/near infrared
SDT	Shock-to-detonation	VD	Variable density
SED	Spectral energy distribution	VIP	Variable inflection point
SF-1	Steroidogenic factor	W	Tungsten
SHAPE	Selective 2' hydroxyl acylation by primer extension	wt	Wild type
SHPB	Split Hopkinson pressure bar	XDT	Anomalous-to-detonation
SIMD	Single instruction, multiple data	XGEL	X-ray free elecdtron laser
SLE	Sea-level equivalent	Xist	X-inactive specific transcript
SMBH	Supermassive black hole	XMM-Newton	X-ray Multi-Mirror Mission-Newton observatory
SN	Discrete ordinate	Zr	Zirconium
Sn	Tin	ZrB_2	Zirconium diboride
SNe	Supernovae	CDM	Lambda cold dark matter
SNP	Single nucleotide polymorphism		
Sos1	Son of sevenless-1		
SRA	Steroid receptor RNA activator		
SRAP	SRA protein		
SSD	Solid state device		
SSE	Streaming SIMD extension		
SST	Sea-surface temperature		
SVE	Statistical volume element		

Codes/Algorithms
ABAQUS
AMR
ARGO
ATOMIC
CAPSAICIN
Cassio
CASTRO
CESM
CFDNS
CG
Cubit
ddcMD
DDMC
FLASH
FRAPCON
GMRES
Hydra-TH
HydroFlow
IMC
INL
IOzone
JFNK
KEPLER
KIVA
LATTE
LIFEIV
lokern
ML
MPAS-Ocean
Multi-fluid PPM
NDA-NCA
NOX
OPLIB

Codes/Algorithms
PARTISN
PIC
POP
PSAAP
PYTHON
QR
RAGE
RDCA
RIID
RMS
SNL
SNSPH
SP2
Spectrum
STRIPACK
Trilinos
TYCHO
VASP
VPSC

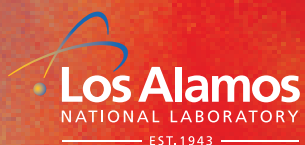| Measurement abbreviations	
A	Atwood number
Å	Ångstrom
cc	Cubic centimeter
cm	Centimeter
eV	Electronvolt
fs	Femtosecond
g	gram
GB	Gigabyte
GPa	Gigapascal
K	Degrees Kelvin
kg	Kilogram
km	Kilometer
km^3	Cubic kilometer
KPa	Kilopascal
m	Meter
M^3	Cubic meter
micron	Micrometer
MPa	Megapascal
ms	Millisecond
nm	Nanometer
ns	Nanosecond
Pr	Prandtl number
ps	picosecond
s	Second
sec	Second
TB	Terabyte
V	Electric potential
μ s	Microsecond
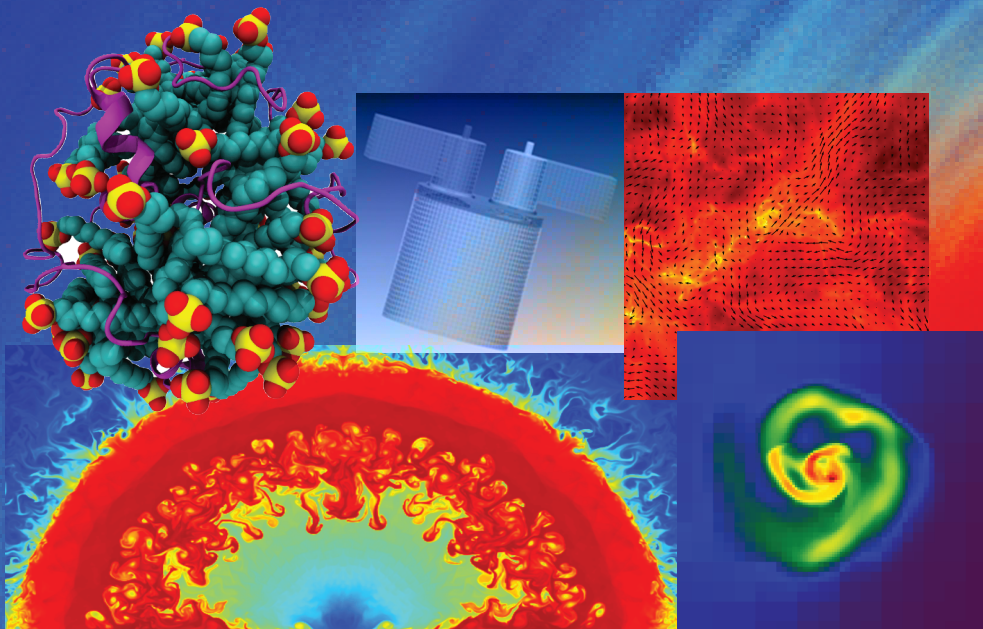
Advanced Computing Solutions
ACS

Computer, Computational, and Statistics Science
CCS Division

High Performance Computing
HPC Division

Theoretical Division
T Division

Los Alamos
NATIONAL LABORATORY
EST. 1943

LA-UR-12-20429